(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 780 472 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.06.1997 Bulletin 1997/26

(21) Application number: 96120662.0

(22) Date of filing: 20.12.1996

(51) Int. Cl.[6]: C12N 15/12, C07K 14/435, C12N 1/21, C12N 15/70, C07K 16/18, A61K 31/70, C12Q 1/68, A61K 39/00, G01N 33/577, C12N 15/79

(84) Designated Contracting States:
AT BE CH DE ES FR GB IT LI NL SE

(30) Priority: 20.12.1995 JP 349661/95
23.07.1996 JP 213181/96

(71) Applicant: Hsp Research Institute, Inc.
Chuo-ku, Osaka-shi, Osaka (JP)

(72) Inventors:
- Ikeda, Jun
  Tokyo (JP)
- Kaneda, Sumiko
  Kyoto-shi, Kyoto (JP)
- Yanagi, Hideki
  Takarazuka-shi, Hyogo-ken (JP)
- Matsumoto, Masayasu
  Mino-shi, Osaka (JP)
- Yura, Takashi
  Kyoto-shi, Kyoto (JP)

(74) Representative: VOSSIUS & PARTNER
Postfach 86 07 67
81634 München (DE)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Stress proteins

(57) Described is a stress protein named ORP150, polynucleotides encoding said protein as well as antibodies against the ORP150 protein. Furthermore, pharmaceutical compositions comprising these proteins, polynucleotides or antibodies are described and their use for the treatment of ischemic diseases.

EP 0 780 472 A2

Printed by Rank Xerox (UK) Business Services
2.14.8/3.4

Description

The present invention relates to an oxygen-regulated protein 150 (ORP150). Specifically, the invention relates to the amino acid sequence of such ORP150 polypeptides, polynucleotides encoding ORP150 polypeptides, promoters of ORP150 genes and antibodies specific to ORP150 polypeptides.

Since the expression of a 70 kDa heat shock protein (HPS70) in cerebral ischemic lesions was reported for the first time, various stress proteins, represented by HSP70, have been reported to be expressed in myocardial ischemic and atherosclerotic lesions, as well as cerebral ischemic lesions. The fact that the induction of HSP, a mechanism of defence against heat stress, is seen in ischemic lesions, suggests that the stress response of the body to ischemic hypoxia is an active phenomenon involving protein neogenesis. Regarding cultured cells, stressful situations that cause ischemia in vivo, such as hypoglycemia and hypoxia, have been shown to induce a group of non-HSP stress proteins, such as glucose-regulated protein (GRP) and oxygen-regulated protein (ORP).

ORP is therefore expected to serve in the diagnosis and treatment of ischemic diseases.

Hori et al. have recently found that exposure of cultured rat astrocytes to hypoxic conditions induces 150, 94, 78, 33 and 28 kDa proteins [J. Neurochem., 66, 973-979(1996)]. These proteins, other than the 150 kDa protein, were identified as GRP94, GRP78, hemoxygenase 1 and HSP28, respectively, while the 150 kDa protein (rat ORP150) remains not to be identified. In addition, there has been no report of human ORP150 protein.

Accordingly, the technical problem underlying the present invention is to provide ORP150 proteins, namely those of human and rat origin, the amino acid sequences of these proteins as well as nucleotide sequences encoding these proteins, the promoter regions of the corresponding genes and antibodies against ORP150 proteins or fragments thereof which are useful in the diagnosis and treatment of ischemic diseases.

This technical problem has been solved by the provision of the embodiments characterized in the claims.

Thus, in a first aspect, the present invention relates to a polynucleotide encoding an ORP150 polypeptide selected from the group consisting of:

(a) polynucleotides encoding the polypeptide having the amino acid sequence as depicted in SEQ ID NO:1 or a fragment of the polypeptide;
(b) polynucleotides comprising the coding region of the nucleotide sequence as shown in SEQ ID NO:2 or a fragment thereof;
(c) polynucleotides encoding the polypeptide having the amino acid sequence as depicted in SEQ ID NO:3 or a fragment of the polypeptide;
(d) polynucleotides comprising the coding region of the nucleotide sequence as depicted in SEQ ID NO:4 or a fragment thereof;
(e) polynucleotides encoding an ORP150 polypeptide which differs from the polypeptide encoded by the polynucleotide of (a) or (c) due to deletion(s), addition(s), insertion(s) and/or substitutions (s) of one or more amino acid residues; and
(f) polynucleotides the complementary strand of which hybridizes to a polynucleotide of any one of (a) to (e) and which encode an ORP150 polypeptide;

and the complementary strand of such a polynucleotide.

In still another embodiment, the present invention relates to a polynucleotide capable of hybridizing to the above polynucleotide or a fragment thereof and having promoter activity.

In still another embodiment, the present invention relates to a recombinant DNA, e.g. vectors, which contains a nucleotide sequence of the present invention.

In still another embodiment, the present invention relates to an expression vector which contains the recombinant DNA of the present invention, to host cells transformed with polynucleotides or vectors of the invention and to a process for the production of an ORP150 protein by cultivating such host cells. In a further embodiment, the present invention relates to the polypeptides encoded by the polynucleotides of the invention.

In still another embodiment, the present invention relates to an antibody or fragment thereof which specifically binds to the polypeptide of the present invention, and to nucleic acid molecules which specifically hybridize to polynucleotides of the present invention.

In still another embodiment the present invention relates to pharmaceutical and diagnostic compositions comprising the above-described polynucleotides, polypeptides, antibodies and/or nucleic acid molecules.

Figure 1 indicates a schematic diagram of the exon-intron structure of the human ORP gene. Black squares represent the exons.

Figure 2 shows the results of the Northern blot analysis of ORP150 mRNA extracted from human astrocytoma U373 cells after exposure to various types of stress.

Figure 3 shows the results of the Northern blot analysis of ORP150 mRNA from adult human tissues.

One embodiment of a polynucleotide of the present invention is a polynucleotide encoding a polypeptide compris-

ing the amino acid sequence shown by SEQ ID NO:1 in the sequence listing, and constituting the human oxygen-regulated protein ORP150 which is obtainable by inducement under hypoxic conditions. Another embodiment of a polynucleotide of the present invention is a polynucleotide encoding a polypeptide comprising the amino acid sequence shown by SEQ ID NO: 3 in the sequence listing, and constituting the rat oxygen-regulated protein ORP150 which is obtainable by inducement under hypoxic conditions. The polynucleotides of the present invention also include those which code for polypeptides each comprising a portion of the above-described polypeptides, and those encoding the entire or portion of the above-described polypeptides. It is a well-known fact that mutation occurs in nature; some of the amino acids of ORP150 protein may be replaced or deleted, and other amino acids may be added or inserted. Mutation can also be induced by gene engineering technology. It is therefore to be understood that substantially homologous polypeptides resulting from such mutations in one or more amino acid residues are also included in the scope of the present invention as long as they are obtainable by inducement under hypoxic conditions.

Further embodiments of a polynucleotide of the present invention are polynucleotides comprising the nucleotide sequence shown by SEQ ID NO:2 in the sequence listing, i.e., human ORP150 cDNA and polynucleotides comprising the nucleotide sequence shown by SEQ ID NO:4 in the sequence listing which represents rat ORP150 cDNA. Polynucleotides comprising a portion of these polynucleotides, and those containing the entire or portion of these polynucleotides are also included in the scope of the present invention. As stated above, the ORP150 gene may have some bases replaced, deleted, added or inserted by mutations, and the resulting polynucleotides with partially different nucleotide sequences are also included in the scope of the present invention, as long as they are substantially homologous and encode a polypeptide obtainable by inducement under hypoxic conditions.

The present invention also relates to a polynucleotide the complementary strand of which hybridizes to a polynucleotide as described above and which codes for an ORP150 polypeptide, this means for a polypeptide inducible under hypoxic conditions. "Hybridizing" in this regard means preferably hybridization under stringent conditions. The hybridizing polynucleotides have preferably a sequence identity of at least 50% most preferably of at least 70%, with the polynucleotides described above. The term "stringent conditions" means that hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

The polynucleotides of the present invention may be RNA or DNA molecules. DNA molecules can, for example, be cDNA, genomic DNA, double or single stranded DNA, isolated from natural sources, produced in vitro or by chemical synthesis methods. The polynucleotides of the invention can code for an ORP150 polypeptide from any organism expressing such a polypeptide, preferably from eukaryots, for example, insects, vertebrates, preferably mammals and most preferably from human, rat, mouse, bovine, sheep, goat or pig.

Furthermore, the present invention also relates to recombinant nucleic acid molecules which comprise a polynucleotide according to the invention. Examples for such molecules are vectors, namely plasmids, cosmids, phagemids, recombinant phages, viruses etc.

In a preferred embodiment the polynucleotide according to the invention present in such a recombinant nucleic acid molecule is linked to regulatory elements which allow for expression in prokaryotic or eukaryotic host cells. Such regulatory elements are well known in the art and include promoters, transcriptional and translational enhancers and the like.

The term "recombinant DNA" as used herein is defined as any DNA containing a polynucleotide described above.

The term "expression vector" as used herein is defined as any vector containing the recombinant DNA of the present invention and expressing a desired protein by introduction into the appropriate host.

The term "clone" as used herein means not only a cell into which a polynucleotide of interest has been introduced but also the polynucleotide of interest itself.

The term "inducement under hypoxic conditions" used herein means an increase in protein synthesis upon exposing cells to an oxygen-depleted atmosphere.

The present invention furthermore relates to host cells transformed and genetically engineered with a polynucleotide according to the invention. These may be prokaryotic or eukaryotic ells. They may be homologous or heterologous with respect to the introduced polynucleotide. If they are homologous they can be distinguished from naturally occurring cells by the feature that they comprise in addition to a naturally occurring ORP150 gene, at least one further copy of an ORP150 coding region which is integrated into the genome in a position in which it does normally not occur. This can be confirmed, e.g., by Southern blotting. Suitable host cells include, for example, bacteria such as E. coli and Bacillus subtilis, yeast such as S. cerevisiae, vertebrate cells, insect cells, mammalian cells, e.g. rat, mouse or human cells.

Moreover, the present invention relates to a process for the production of an ORP150 polypeptide which comprises the steps of culturing the host according to the invention and recovering the produced polypeptide from the cells and/or the culture medium.

The present invention also relates to the polypeptides encoded by the polynucleotides according to the invention or obtainable by the above described process.

The amino acid sequences and nucleotide sequences of the present invention can, for example, be determined as follows: First, poly(A)$^+$ RNA is prepared from rat astrocytes exposed to hypoxic conditions. After cDNA is synthesized from said poly(A)$^+$RNA using random hexamer primers, a cDNA library is prepared using the pSPORT1 vector (pro-

duced by Life Technology), or the like.

Next, PCR is conducted using oligonucleotide primers synthesized on the basis of the nucleotide sequence of the pSPORT1 vector used to prepare the cDNA library above and the degenerate nucleotide sequences deduced from the N-terminal amino acid sequence of purified rat ORP150, to yield a large number of amplified DNA fragments. These DNA fragments are then inserted into the pT7 Blue vector (produced by Novagen), or the like, for cloning to obtain a clone having nucleotide sequence which perfectly encodes the N-terminal amino acid sequence. Purification of ORP150 can be achieved by commonly used methods of protein purification, such as column chromatography and electrophoresis, in combination as appropriate.

In addition, by screening the above-described rat astrocyte cDNA library by colony hybridization using the insert in above clone as a probe, a clone having an insert thought to encode rat ORP150 can be obtained. This clone is subjected to stepwise deletion from both the 5'- and 3'-ends, and oligonucleotide primers prepared from determined nucleotide sequences are used to determine the nucleotide sequence sequentially. If the clone thus obtained does not encode the full length of rat ORP150, an oligonucleotide probe is synthesized on the basis of the nucleotide sequence of the 5'- or 3'-region of the insert, followed by screening for a clone containing the nucleotide sequence extended further in the 5' or 3' direction, for example, the Gene Trapper cDNA Positive Selection System Kit (produced by Life Technology) based on hybridization using magnetic beads. The full-length cDNA of the rat ORP150 gene is thus obtained.

Separately, the following procedure is followed to obtain a human homologue of rat ORP150 cDNA. Poly(A)$^+$RNA is prepared from the human astrocytoma U373 exposed to hypoxic conditions. After cDNA is synthesized from said poly(A)$^+$RNA using random hexamer primers and an oligo(dT) primer, said cDNA is inserted into the EcoRI site of the pSPORT1 vector to prepare a cDNA library. Human ORP150 cDNA is then obtained using the Gene Trapper Kit and the nucleotide sequence is determined in the same manner as with rat ORP150 above.

The nucleotide sequence of human ORP150 cDNA is thus determined as that shown by SEQ ID NO:2 in the sequence listing, based on which the amino acid sequence of human ORP150 is determined.

Exposure of astrocytes to hypoxic conditions can, for example, be achieved by the method of Ogawa et al. [Ogawa, S., Gerlach, H., Esposito, C., Mucaulay, A.P., Brett, J., and Stern, D., J. Clin. Invest., 85, 1090-1098 (1990)].

Furthermore, the following procedure is followed to obtain human ORP150 genomic DNA. A genomic library purchased from Clontech (derived from human placenta, Cat. #HL1067J) is used. Screening is conducted by hybridization using a DNA fragment consisting of 202 bp of the 5' untranslated region and 369 bp of the coding region, derived from the rat cDNA clone, as well as a 1351 bp DNA fragment containing the termination codon, derived from the human cDNA, as probes. Two clones containing the ORP150 gene are isolated, one containing exons 1 through 24 and the other containing exons 16 through 26; the entire ORP150 gene is composed by combining these two clones. The nucleotide sequence of the 15851 bp human ORP150 genomic DNA is determined; its nucleotide sequence from the 5'-end to just before the translation initiation codon ATG in exon 2 is shown by SEQ ID NO:12 in the sequence listing.

As stated above, the present invention includes polypeptides containing the entire or portion of the polypeptide (human ORP150) having the amino acid sequence shown by SEQ ID NO:1 in the sequence listing. The present invention also includes the entire or portion of the polypeptide having the amino acid sequence shown by SEQ ID NO:1 in the sequence listing; for example, polynucleotides containing the entire or portion of the nucleotide sequence shown by SEQ ID NO:2 in the sequence listing are included in the scope of the present invention. The present invention also includes specific antibodies against these polypeptides of the present invention, and fragments thereof.

An antibody against a polypeptide of the present invention, which polypeptide contains the entire or portion of human or rat ORP150, can be prepared by a conventional method [Current Protocols in Immunology, Coligan, J.E. et al. eds., 2.4.1-2.4.7, John Wiley & Sons, New York (1991)]. Specifically, a rat ORP150 band, separated by, for example, SDS-polyacrylamide gel electrophoresis, is cut out and given to a rabbit etc. for immunization, after which blood is collected from the immunized animal to obtain an antiserum. An IgG fraction can be obtained if necessary by affinity chromatography using immobilized protein A, or the like. A peptide identical to the partial amino acid sequence of ORP150 can be chemically synthesized as a multiple antigen peptide (MAP) [Tam, J.P., Proc. Natl. Acad. Sci. USA, 85, 5409-5413 (1988)], and can be used for immunization in the same manner as above.

It is also possible to prepare a monoclonal antibody by a conventional method [Cell & Tissue Culture; Laboratory Procedure (Doyle, A. et al., eds.) 25A:1-25C:4, John Wiley & Sons, New York (1994)] using a polypeptide containing the entire or portion of human or rat ORP150 as an antigen. Specifically, a hybridoma is prepared by fusing mouse splenocytes immunized with said antigen and a myeloma cell line, and the resulting hybridoma is cultured or intraperitoneally transplanted to the mouse to produce a monoclonal antibody.

The fragments resulting from protease digestion of these antibodies as purified can also be used as antibodies of the present invention.

The present invention also relates to nucleic acid molecules which specifically hybridize with a polynucleotide according to the invention or with the complementary strand of such a polynucleotide. "Specifically hybridizing" means that such molecules show no significant cross-hybridization to polynucleotides coding for proteins other than an ORP150 polypeptide. Preferably these nucleic acid molecules have a length of at least 15 nucleotides, more preferably of at least 30 nucleotides and most preferably of at least 50 nucleotides. In a preferred embodiment these molecules

have over their entire length a sequence identity to a corresponding region of a polynucleotide of the invention of at least 85%, preferably of at least 90% and most preferably of at least 95%. In a particularly preferred embodiment the sequence identity is at least 97%. These nucleic acid molecules can be used, for example, as hybridization probes for the isolation of related genes, as PCR primers, for the diagnosis of mutations of ORP150 genes, for the use in antisense molecules or ribozymes or the like.

The polynucleotides of the present invention, the polypeptides encoded by them, specific antibodies against these polypeptides or fragments thereof and the nucleic acid molecules specifically hybridizing to the above-mentioned polynucleotides are useful in the diagnosis and treatment of ischemic diseases, permitting utilization for the development of therapeutic drugs for ischemic diseases.

Thus, the present invention also relates to a pharmaceutical composition comprising a polynucleotide, polypeptide, antibody and/or nucleic acid molecule according to the invention. Optionally, such a composition also comprises a pharmaceutically acceptable carrier.

The invention also relates to diagnostic composition comprising a polynucleotide, polypeptide, antibody and/or nucleic acid molecule according to the invention.

In another embodiment the present invention relates to a polynucleotide comprising or containing the entire or portion of the nucleotide sequence shown by SEQ ID NO:12 in the sequence listing. This is a polynucleotide containing the promoter region of the human ORP150 gene. Polynucleotides capable of hybridizing to this polynucleotide under conventional hybridizing conditions (e.g., in 0.1 x SSC containing 0.1% SDS at 65°C) and possessing promoter activity are also included in the scope of the present invention. Preferably, such a promoter is able to promote transcription in cells when exposed to hypoxia. Successful cloning of said promoter region would dramatically advance the functional analysis of the human ORP150 gene and facilitate its application to the treatment of ischemic diseases.

The term "promoter" as used herein is defined as a polynucleotide comprising a nucleotide sequence that activates or suppresses the transcription of a desired gene by being present upstream or downstream of said gene.

The following examples illustrate the present invention

Example 1

Cell culture and achievement of hypoxic condition

Rat primary astrocytes and microglia were obtained from neonatal rats by a modification of a previously described method [Maeda, Y., Matsumoto, M., Ohtsuki, T., Kuwabara, K., Ogawa, S., Hori, O., Shui, D.Y., Kinoshita, T., Kamada, T., and Stern, D., J. Exp. Med., 180, 2297-2308(1994)]. Briefly, cerebral hemispheres were harvested from neonatal Sprague-Dawley rats within 24 hours after birth, meninges were carefully removed, and brain tissue was digested at 37°C in minimal essential medium (MEM) with Joklik's modification (Gibco, Boston MA) containing Dispase II (3mg/ml; Boehringer-Mannheim, Germany). After centrifugation, the cell pellet was resuspended and grown in MEM supplemented with fetal calf serum (FCS; 10%; CellGrow, MA).

After 10 days, cytosine arabinofuranoside (10μg/ml; Wako Chemicals, Osaka, Japan) was added for 48 hours to prevent fibroblast overgrowth, and culture flasks were agitated on a shaking platform. Then, floating cells were aspirated (these were microglia), and the adherent cell population was identified by morphological criteria and immunohistochemical staining with anti-glial fibrillary acidic protein antibody. Cultures used for experiments were >98% astrocytes based on these techniques.

Human astrocytoma cell line U373 was obtained from American Type Culture Collection (ATCC) and cultured in Dulbecco's modified Eagle medium (produced by Life Technology) supplemented with 10% FCS.

Cells were plated at a density of about 5 X 10^4 cells /cm^2 in the above medium. When cultures achieved confluence, they were exposed to hypoxia using an incubator attached to a hypoxia chamber which maintained a humidified atmosphere with low oxygen tension (Coy Laboratory Products, Ann Arbor MI) as described previously [Ogawa, S., Gerlach, H., Esposito, C., Macaulay, A.P., Brett, J., and Stern, D., J. Clin. Invest., 85, 1090-1098 (1990)].

Example 2

Purification and N-terminal sequencing of the rat 150 kDa polypeptide

Rat primary astrocytes (about 5 x 10^8 cells) exposed to hypoxia for 48 hours were harvested, cells were washed three times with PBS(pH 7.0) and protein was extracted with PBS containing NP-40 (1%), PMSF (1mM), and EDTA (5mM). Extracts were then filtered (0.45 μm nitrocellulose membrane), and either subjected to reduced SDS-PAGE (7.5%, about 25μg) or 2-3 mg of protein was diluted with 50 ml of PBS (pH 7.0) containing NP-40(0.05%) and EDTA (5mM), and applied to FPLC Mono Q(bed volume 5 ml, Pharmacia, Sweden).

The column was washed with 0,2M NaCl, eluted with an ascending salt gradient (0.2 to 1.8 M NaCl) and 10 μl of each fraction (0.5 ml) was applied to reduced SDS-PAGE (7.5%), along with molecular weight markers (Biorad). Pro-

teins in the gel were visualized by silver staining. Fractions eluted from FPLC Mono Q which contained the 150 kDa polypeptide (#7-8) were pooled and concentrated by ultrafiltration (Amicon) 50-fold and about 200 μg of protein was applied to preparative, reduced SDS-PAGE (7.5%). Following electrophoresis, proteins in the gel were transferred electrophoretically ($2A/cm^2$) to polyvinylidene difluoride (PVDF) paper (Millipore, Tokyo), the paper was dried, stained with Coomassie Brilliant blue, and the band corresponding to 150 kDa protein (OPR150) was cut out for N-terminal sequencing using an automated peptide sequencing system (Applied Biosystems, Perkin-Elmer). The N-terminal 31-amino acid sequence was thus determined (SEQ ID NO:5).

Example 3

Preparation of rat astrocyte cDNA library

Total RNA was prepared from rat primary astrocytes (1.1×10^8 cells), in which ORP150 had been induced under hypoxic conditions, by the acid guanidinium-phenol-chloroform method [Chomczynski, P. and Sacchi, N., Anal. Biochem., 162, 156-159 (1987)]. Using 300 μg of the total RNA obtained, purification was conducted twice in accordance with the protocol for poly(A)$^+$ RNA purification using oligo(dT)-magnetic beads (produced by Perceptive Diagnostics), to yield poly(A)$^+$ RNA. Double-stranded cDNA was then synthesized using random hexamer primers, in accordance with the protocol for the Superscript Choice System (produced by Life Technology), and inserted into the EcoRI site of the pSPORT1 vector to prepare a cDNA library consisting of 5.4×10^5 independent clones.

Example 4

Cloning of rat ORP150 cDNA

Rat ORP150 cDNA was cloned as follows: First, to obtain a probe for colony hybridization, the cDNA library was subjected to PCR using a 20-base primer, 5'-AATACGACTCACTATAGGGA-3' (SEQ ID NO:6), which corresponds to the antisense strand of the T7 promoter region in the pSPORT1 vector, and 20 base mixed primers, 5'-AARCCiGGiGTNCCNATGGA-3' (SEQ ID NO:8), which contains inosine residues and degenerate polynucleotides and which was prepared on the basis of the oligonucleotide sequence deduced from a partial sequence (KPGVPME) (SEQ ID NO:7) within the N-terminal amino acid sequence (LAVMSVDLGSESMKVAIVKPGVPMEIVLNKE) (SEQ ID NO:5); the resulting PCR product with a length of about 480 bp was inserted into the pT7 Blue Plasmid vector. Nucleotide sequences of the clones containing an insert of the expected size (480 bp) corresponding to the PCR product were determined using an automatic nucleotide sequencer (produced by Perkin-Elmer, Applied Biosystems). A clone containing a 39-nucleotide sequence encoding a peptide identical to the rat ORP150-specific amino acid sequence KPGVPMEIVLNKE (SEQ ID NO:9) in the insert was thus obtained.

Using the above insert of the clone as a probe, RNA from cultured rat astrocytes were subjected to Northern blotting; the results demonstrated that mRNA with a length of about 4 Kb was induced by hypoxic treatment. Thereupon, the above insert of the clone was labeled by the random prime labeling method (Ready TOGO, produced by Pharmacia) using α-[^{32}P]dCTP to yield a probe. Using this probe, 1.2×10^4 clones of the cDNA library were screened by colony hybridization to obtain a clone containing a 2800 bp insert. The nucleotide sequence of this clone insert was determined by preparing deletion mutants using a kilosequence deletion kit (produced by Takara Shuzo).

Since this clone did not contain the 3'-region of the ORP150 coding sequence, the following two 20-base oligonucleotides were prepared on the basis of the specific nucleotide sequence near the 3' end of the above insert, to obtain the full-length sequence.

5'-GCACCCTTGAGGAAAATGCT-3' (SEQ ID NO:10)
5'-CCCAGAAGCCCAATGAGAAG-3' (SEQ ID NO:11)

Using the two oligonucleotides, a clone containing the entire coding region was selected from the rat astrocyte cDNA library in accordance with the protocol for the Gene Trapper cDNA Positive Selection System (produced by Life Technology), and its nucleotide sequence was determined.

The nucleotide sequence of rat ORP150 cDNA was thus determined as shown by SEQ ID NO:4 in the sequence listing. Based on this nucleotide sequence, the amino acid sequence of rat ORP150 was determined as shown by SEQ ID NO:3 in the sequence listing.

Example 5

Preparation of human U373 cDNA library

Poly(A)$^+$ RNA was purified from U373 cells (1×10^8 cells) in which human ORP150 had been induced under hypoxic conditions, in the same manner as described in Example 3. Double-stranded cDNA was then synthesized in

accordance with the protocol for the Superscript Choice System (produced by Life Technology) using a 1:1 mixture of random hexamer primers and an oligo(dT) primer. This cDNA was inserted into the EcoRI site of the pSPORT1 vector to prepare a cDNA library consisting of 2×10^5 independent clones.

Specifically, the library was prepared as follows: Human U373 cells, cultured in 10 plastic petri dishes (150 mm in diameter)(1×10^7 cells/dish), were subjected to hypoxic treatment for 48 hours by the method of Ogawa et al. [Ogawa, S., Gerlach, H., Esposito, C., Mucaulay, A.P., Brett, J., and Stern, D., J. Clin. Invest., 85, 1090-1098 (1990)] as described in Example 3, after which total RNA was prepared by the acid guanidinium-phenol-chloroform method [Chomczynski, P. and Sacchi, N., Anal. Biochem., 162, 156-159 (1987)]. Using 500 μg of the total RNA obtained, purification was conducted twice in accordance with the protocol for poly(A)$^+$ RNA purification using oligo(dT)-magnetic beads (produced by Perceptive Diagnostics), to yield poly(A)$^+$ RNA. Double-stranded cDNA was then synthesized using 5 μg of the poly(A)$^+$ RNA and a 1:1 mixture of random hexamer primers and an oligo(dT) primer, in accordance with the protocol for the Superscript Choice System (produced by Life Technology), and inserted into the EcoRI site of the pSPORT1 vector to prepare a human U373 cDNA library consisting of 2×10^5 independent clones.

Example 6

Cloning of human ORP150 cDNA

Using two primers (SEQ ID NO:10 and SEQ ID NO:11) prepared on the basis of the above-described rat ORP150 cDNA specific sequence, a clone containing the entire coding region was selected from the human U373 cDNA library in accordance with the protocol for the Gene Trapper cDNA Positive Selection System (produced by Life Technology), and its nucleotide sequence was determined. The nucleotide sequence of human ORP150 cDNA was thus determined as shown by SEQ ID NO:2 in the sequence listing.

Specifically, 2×10^4 clones of the human U373 cDNA library were amplified in accordance with the protocol for the Gene Trapper cDNA Positive Selection System (produced by Life Technology). Five micrograms of the plasmid purified from amplified clones were treated with the Gene II and Exo III nuclease included in the kit to yield single-stranded DNA. An oligonucleotide (SEQ ID NO:10) prepared on the basis of the above-described rat ORP150 cDNA-specific sequence was biotinylated and subsequently hybridized to the above single-stranded DNA at 37°C for 1 hour. The single-stranded DNA hybridized to the oligonucleotide derived from rat ORP150 cDNA was selectively recovered by using streptoavidin-magnetic beads, and was treated with the repair enzyme included in the kit using the oligonucleotide shown by SEQ ID NO:10 in the sequence listing as a primer, to yield double-stranded plasmid DNA.

The double-stranded plasmid DNA was then introduced to ElectroMax DH10B cells (produced by Life Technology) in accordance with the protocol for the Gene Trapper cDNA Positive Selection System, followed by colony PCR in accordance with the same protocol using two primers (SEQ ID NO:10 and SEQ ID NO:11) prepared on the basis of the rat ORP150 cDNA-specific sequence, to select clones that yield an about 550 bp PCR product. The nucleotide sequence of the longest insert among these clones, corresponding to the human ORP150 cDNA, was determined as shown by SEQ ID NO:2 in the sequence listing.

On the basis of this nucleotide sequence, the amino acid sequence of human ORP150 was determined as shown by SEQ ID NO:1 in the sequence listing.

The N-terminal amino acid sequence (SEQ ID NO: 5) obtained with purified rat ORP150 corresponded to amino acids 33-63 deduced from both the human and rat cDNAs, indicating that the first 32 residues represent the signal peptides for secretion. The C-terminal KNDEL sequence, which resembles KDEL sequence, a signal to retain the ER-resident proteins [Pelham, H.R.B., Trends Biochem. Sci. 15, 483-486 (1990)], may function as an ER-retention signal. The existence of a signal peptide at the N-terminus and the ER-retention signal-like sequence at the C-terminus suggests that ORP150 resides in the ER, consistent with the results of immunocytochemical analysis reported by Kuwabara et al. [Kuwabara, K., Matsumoto, M., Ikeda, J., Hori, O., Ogawa, S., Maeda, Y., Kitagawa, K., Imuta, N., Kinoshita, T., Stern, D.M., Yanagi, H., and Kamada, T., J. Biol. Chem. 271, 5025-5032 (1996)].

Analysis of protein data bases with the BLAST program [Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., J. Mol., Biol. 215, 403-410(1990)] showed that the N-terminal half of ORP150 has a modest similarity to the ATPase domain of numerous HSP70 family sequences. An extensive analysis with pairwise alignments [Pearson, W.R., and Lipman, D.J., Proc. Natl. Acad. Sci. USA 85, 2444-2448(1988)] revealed that amino acids 33-426 of human ORP150 was 32% identical to amino acids 1-380 of both inducible human HSP70.1 [Hunt, C., and Morimoto, R.I., Proc. Natl. Acad. Sci. USA 82, 6455-6459 (1985)] and constitutive bovine HSC70 [DeLuca-Flaherty, C., and McKay, D.B., Nucleic Acids Res. 18, 5569(1990)], typical members of HSP70 family. An additional region similar to HSP70RY and hamster HSP110, which both belong to a new subfamily of large HSP70-like proteins [Lee-Yoon, D., Easton, D., Murawski, M., Burd, R., and Subjeck, J.R., J. Biol. Chem. 270, 15725-15733 (1995)], extended further to residue 487. A protein sequence motif search with PROSITE [Bairoch, A., and Bucher, P., Nucleic Acids Res. 22, 3583-3589(1994)] showed that ORP150 contains two of the three HSP70 protein family signatures: FYDMGSGSTVCTIV (amino acids 230-243, SEQ ID NO:1) and VILVGGATRVPRVQE (amino acids 380-394, SEQ ID NO:1) which completely matched

with the HSP70 signatures 2 and 3, respectively, and VDLG (amino acids 38-41, SEQ ID NO:1) which matched with the first four amino acids of the signature 1. Furthermore, the N-terminal region of ORP150 contained a putative ATP-binding site consisting of the regions (amino acids 36-53, 197-214, 229-243, 378-400, and 411-425, SEQ ID NO:1) corresponding to the five motifs specified by Bork et al. [Bork, P., Sander, C., and Valencia, A., Proc. Natl. Acad. Sci. USA 89, 7290-7294 (1992)]. Although the C-terminal putative peptide-binding domains of HSP70 family are generally less conserved [Rippmann, F., Taylor, W.R., Rothbard, J.B., and Green, N.M., EMBO J. 10, 1053-1059 (1991)], the C-terminal region flanked by amino acids 701 and 898 (SEQ ID NO:1) shared appreciable similarity with HSP110 (amino acids 595-793; 29% identity).

Example 7

Cloning of human ORP150 genomic DNA

A human genomic library purchased from Clontech (derived from human placenta, Cat. #HL1067J, Lot #1221, 2.5 x 10^6 independent clones) was used. A DNA fragment consisting of 202 bp of the 5' untranslated region and 369 bp of the coding region derived from the rat cDNA clone, as well as a 1351 bp DNA fragment containing the termination codon, derived from the human cDNA, were used as probes for plaque hybridization.

Escherichia coli LE392, previously infected with 1 x 10^6 pfu of the human genomic library, was plated onto 10 petri dishes 15 cm in diameter to allow plaque formation. The phage DNA was transferred to a nylon membrane (Hybond-N^+, Amersham) and denatured with sodium hydroxide, after which it was fixed by ultraviolet irradiation. The rat cDNA probe was labeled using a DNA labeling kit (Ready To Go, Pharmacia), and hybridized with the membrane in the Rapid-hyb buffer (Amersham). After incubation at 65°C for 2 hours, the nylon membrane was washed with 0.2 x SSC-0.1% SDS, and a positive clone was detected on an imaging plate (Fuji Photo Film). Since the clone isolated contained only exons 1 through 24, 1.5 x 10^6 clones of the same library was screened again using the human cDNA probe in the same manner, resulting in isolation of one clone. This clone was found to contain exons 16 through 26, with an overlap with the 3' region of the above-mentioned clone. The entire region of the ORP150 gene was thus cloned by combining these two clones.

These two clones were cleaved with BamHI and subcloned into pBluescript IISK (Stratagene), followed by nucleotide sequence determination of the entire 15851 bp human ORP150 genomic DNA. The nucleotide sequence from the 5' end to just before the translation initiation codon ATG in exon 2 is shown by SEQ ID NO:12 in the sequence listing.

Furthermore, the nucleotide sequence of the 15851 bp human ORP150 genomic DNA was compared with that of the human ORP150 cDNA shown by SEQ ID NO:2 in the sequence listing, resulting in the demonstration of the presence of the exons at the positions shown below. A schematic diagram of the positions of the exons is shown in Figure 1.

		(Base position in SEQ ID:2)
Exon 1	1908 - 2002	(1 - 95)
Exon 2	2855 - 2952	(96 - 193)
Exon 3	3179 - 3272	(194 - 287)
Exon 4	3451 - 3529	(288 - 366)
Exon 5	3683 - 3837	(367 - 521)
Exon 6	3962 - 4038	(522 - 598)
Exon 7	4347 - 4528	(599 - 780)
Exon 8	4786 - 4901	(781 - 896)
Exon 9	6193 - 6385	(897 - 1089)
Exon 10	6593 - 6727	(1090 - 1224)
Exon 11	6850 - 6932	(1225 - 1307)
Exon 12	7071 - 7203	(1308 - 1440)
Exon 13	7397 - 7584	(1441 - 1628)
Exon 14	7849 - 7987	(1629 - 1767)
Exon 15	9176 - 9236	(1768 - 1828)
Exon 16	9378 - 9457	(1829 - 1908)
Exon 17	9810 - 9995	(1909 - 2094)
Exon 18	10127 -10299	(2095 - 2267)
Exon 19	10450 -10537	(2268 - 2355)
Exon 20	10643 -10765	(2356 - 2478)
Exon 21	10933 -11066	(2479 - 2612)
Exon 22	11195 -11279	(2613 - 2697)
Exon 23	12211 -12451	(2698 - 2938)
Exon 24	12546 -12596	(2939 - 2989)
Exon 25	13181 -13231	(2990 - 3040)
Exon 26	13358 -14823	(3041 - 4503)

Example 8

Northern blot analysis

A 4.5-kb EcoRI fragment of human ORP150 cDNA was labeled with [α-^{32}P]dCTP(3,000 Ci/mmol; Amersham Corp., Arlington Heights, IL) by using a DNA labeling kit (Pharmacia), and used as a hybridization probe. 20μg of total RNA prepared from U373 cells exposed to various stresses were electrophoresed and transferred onto a Hybond N$^+$ membrane (Amersham Corp.). Multiple Tissue Northern Blots, in which each lane contained 2μg of poly(A)RNA from the adult human tissues indicated, was purchased from Clontech. The filter was hybridized at 65°C in the Rapid-hyb buffer (Amersham Corp.) with human ORP150, GRP78, HSP70, glyceraldehyde-3-phosphate dehydrogenase (G3PDH), and β-actin cDNAs each labeled with [α^{32}-P] dCTP, washed with 0.1 x SSC containing 0.1% SDS at 65°C, and followed by autoradiography.

As shown in Figure 2, the ORP150 mRNA level was highly enhanced upon 24 - 48 hours of exposure to hypoxia. In parallel experiments, treatment with 2-deoxyglucose (25 mM, 24 hours) or tunicamycin (5μg/ml, 24 hours) enhanced

ORP150 mRNA to the levels comparable to that induced by hypoxia. The induction levels were also comparable with those observed for mRNA of a typical glucose-regulated protein GRP78. Heat shock treatment failed to enhance ORP150 mRNA appreciably.

ORP150 mRNA was found to be highly expressed in the liver and pancreas, whereas little expression was observed in kidney and brain (Figure 3). Furthermore, the tissue specificity of ORP150 expression was quite similar to that of GRP78. The higher expression observed in the tissues that contain well-developed ER and synthesize large amounts of secretory proteins is consistent with the finding that ORP150 is localized in the ER (Kuwabara, K., Matsumoto, M., Ikeda, J., Hori, O., Ogawa, S., Maeda, Y., Kitagawa, K., Imuta, N., Kinoshita, T., Stern, D.M., Yanagi, H., and Kamada, T., J. Biol. Chem. 271, 5025-5032(1996)).

In conclusion, both the characteristic primary protein structure and the similarity found with GRP78 in stress inducibility and tissue specificity suggest that ORP150 plays an important role in protein folding and secretion in the ER, perhaps as a molecular chaperone, in concert with other GRPs to cope with environmental stress.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the present invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:
(iii) NUMBER OF SEQUENCES: 12

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Ala Asp Lys Val Arg Arg Gln Arg Pro Arg Arg Arg Val Cys Trp
5 10 15
Ala Leu Val Ala Val Leu Leu Ala Asp Leu Leu Ala Leu Ser Asp Thr
20 25 30
Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
35 40 45
Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu Ser
50 55 60
Arg Arg Lys Thr Pro Val Ile Val Thr Leu Lys Glu Asn Glu Arg Phe
65 70 75 80
Phe Gly Asp Ser Ala Ala Ser Met Ala Ile Lys Asn Pro Lys Ala Thr
85 90 95
Leu Arg Tyr Phe Gln His Leu Leu Gly Lys Gln Ala Asp Asn Pro His
100 105 110
Val Ala Leu Tyr Gln Ala Arg Phe Pro Glu His Glu Leu Thr Phe Asp
115 120 125
Pro Gln Arg Gln Thr Val His Phe Gln Ile Ser Ser Gln Leu Gln Phe
130 135 140
Ser Pro Glu Glu Val Leu Gly Met Val Leu Asn Tyr Ser Arg Ser Leu
145 150 155 160
Ala Glu Asp Phe Ala Glu Gln Pro Ile Lys Asp Ala Val Ile Thr Val
165 170 175
Pro Val Phe Phe Asn Gln Ala Glu Arg Arg Ala Val Leu Gln Ala Ala
180 185 190
Arg Met Ala Gly Leu Lys Val Leu Gln Leu Ile Asn Asp Asn Thr Ala
195 200 205
Thr Ala Leu Ser Tyr Gly Val Phe Arg Arg Lys Asp Ile Asn Thr Thr
210 215 220
Ala Gln Asn Ile Met Phe Tyr Asp Met Gly Ser Gly Ser Thr Val Cys
225 230 235 240
Thr Ile Val Thr Tyr Gln Met Val Lys Thr Lys Glu Ala Gly Met Gln
245 250 255
Pro Gln Leu Gln Ile Arg Gly Val Gly Phe Asp Arg Thr Leu Gly Gly
260 265 270
Leu Glu Met Glu Leu Arg Leu Arg Glu Arg Leu Ala Gly Leu Phe Asn
275 280 285
Glu Gln Arg Lys Gly Gln Arg Ala Lys Asp Val Arg Glu Asn Pro Arg
290 295 300
Ala Met Ala Lys Leu Leu Arg Glu Ala Asn Arg Leu Lys Thr Val Leu
305 310 315 320

```
Ser Ala Asn Ala Asp His Met Ala Gln Ile Glu Gly Leu Met Asp Asp
                325                 330                 335
Val Asp Phe Lys Ala Lys Val Thr Arg Val Glu Phe Glu Glu Leu Cys
            340                 345                 350
Ala Asp Leu Phe Glu Arg Val Pro Gly Pro Val Gln Gln Ala Leu Gln
        355                 360                 365
Ser Ala Glu Met Ser Leu Asp Glu Ile Glu Gln Val Ile Leu Val Gly
    370                 375                 380
Gly Ala Thr Arg Val Pro Arg Val Gln Glu Val Leu Leu Lys Ala Val
385                 390                 395                 400
Gly Lys Glu Glu Leu Gly Lys Asn Ile Asn Ala Asp Glu Ala Ala Ala
                405                 410                 415
Met Gly Ala Val Tyr Gln Ala Ala Ala Leu Ser Lys Ala Phe Lys Val
            420                 425                 430
Lys Pro Phe Val Val Arg Asp Ala Val Val Tyr Pro Ile Leu Val Glu
        435                 440                 445
Phe Thr Arg Glu Val Glu Glu Glu Pro Gly Ile His Ser Leu Lys His
    450                 455                 460
Asn Lys Arg Val Leu Phe Ser Arg Met Gly Pro Tyr Pro Gln Arg Lys
465                 470                 475                 480
Val Ile Thr Phe Asn Arg Tyr Ser His Asp Phe Asn Phe His Ile Asn
                485                 490                 495
Tyr Gly Asp Leu Gly Phe Leu Gly Pro Glu Asp Leu Arg Val Phe Gly
            500                 505                 510
Ser Gln Asn Leu Thr Thr Val Lys Leu Lys Gly Val Gly Asp Ser Phe
        515                 520                 525
Lys Lys Tyr Pro Asp Tyr Glu Ser Lys Gly Ile Lys Ala His Phe Asn
    530                 535                 540
Leu Asp Glu Ser Gly Val Leu Ser Leu Asp Arg Val Glu Ser Val Phe
545                 550                 555                 560
Glu Thr Leu Val Glu Asp Ser Ala Glu Glu Glu Ser Thr Leu Thr Lys
                565                 570                 575
Leu Gly Asn Thr Ile Ser Ser Leu Phe Gly Gly Gly Thr Thr Pro Asp
            580                 585                 590
Ala Lys Glu Asn Gly Thr Asp Thr Val Gln Glu Glu Glu Glu Ser Pro
        595                 600                 605
Ala Glu Gly Ser Lys Asp Glu Pro Gly Glu Gln Val Glu Leu Lys Glu
    610                 615                 620
Glu Ala Glu Ala Pro Val Glu Asp Gly Ser Gln Pro Pro Pro Pro Glu
625                 630                 635                 640
Pro Lys Gly Asp Ala Thr Pro Glu Gly Glu Lys Ala Thr Glu Lys Glu
                645                 650                 655
Asn Gly Asp Lys Ser Glu Ala Gln Lys Pro Ser Glu Lys Ala Glu Ala
            660                 665                 670
Gly Pro Glu Gly Val Ala Pro Ala Pro Glu Gly Glu Lys Lys Gln Lys
        675                 680                 685
Pro Ala Arg Lys Arg Arg Met Val Glu Glu Ile Gly Val Glu Leu Val
    690                 695                 700
Val Leu Asp Leu Pro Asp Leu Pro Glu Asp Lys Leu Ala Gln Ser Val
705                 710                 715                 720
Gln Lys Leu Gln Asp Leu Thr Leu Arg Asp Leu Glu Lys Gln Glu Arg
                725                 730                 735
Glu Lys Ala Ala Asn Ser Leu Glu Ala Phe Ile Phe Glu Thr Gln Asp
            740                 745                 750
Lys Leu Tyr Gln Pro Glu Tyr Gln Glu Val Ser Thr Glu Glu Gln Arg
        755                 760                 765
```

```
Glu Glu Ile Ser Gly Lys Leu Ser Ala Ala Ser Thr Trp Leu Glu Asp
    770                 775                 780
Glu Gly Val Gly Ala Thr Thr Val Met Leu Lys Glu Lys Leu Ala Glu
785                 790                 795                 800
Leu Arg Lys Leu Cys Gln Gly Leu Phe Phe Arg Val Glu Glu Arg Lys
                805                 810                 815
Lys Trp Pro Glu Arg Leu Ser Ala Leu Asp Asn Leu Leu Asn His Ser
            820                 825                 830
Ser Met Phe Leu Lys Gly Ala Arg Leu Ile Pro Glu Met Asp Gln Ile
        835                 840                 845
Phe Thr Glu Val Glu Met Thr Thr Leu Glu Lys Val Ile Asn Glu Thr
    850                 855                 860
Trp Ala Trp Lys Asn Ala Thr Leu Ala Glu Gln Ala Lys Leu Pro Ala
865                 870                 875                 880
Thr Glu Lys Pro Val Leu Leu Ser Lys Asp Ile Glu Ala Lys Met Met
                885                 890                 895
Ala Leu Asp Arg Glu Val Gln Tyr Leu Leu Asn Lys Ala Lys Phe Thr
            900                 905                 910
Lys Pro Arg Pro Arg Pro Lys Asp Lys Asn Gly Thr Arg Ala Glu Pro
        915                 920                 925
Pro Leu Asn Ala Ser Ala Ser Asp Gln Gly Glu Lys Val Ile Pro Pro
    930                 935                 940
Ala Gly Gln Thr Glu Asp Ala Glu Pro Ile Ser Glu Pro Glu Lys Val
945                 950                 955                 960
Glu Thr Gly Ser Glu Pro Gly Asp Thr Glu Pro Leu Glu Leu Gly Gly
                965                 970                 975
Pro Gly Ala Glu Pro Glu Gln Lys Glu Gln Ser Thr Gly Gln Lys Arg
            980                 985                 990
Pro Leu Lys Asn Asp Glu Leu
        995
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE
(A) NAME/KEY: CDS
(B) IDENTIFICATION METHOD: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
TTGTGAAGGG CGCGGGTGGG GGGCGCTGCC GGCCTCGTGG GTACGTTCGT GCCGCGTCTG   60
TCCCAGAGCT GGGGCCGCAG GAGCGGAGGC AAGAGGGGCA CTATGGCAGA CAAAGTTAGG  120
AGGCAGAGGC CGAGGAGGCG AGTCTGTTGG GCCTTGGTGG CTGTGCTCTT GGCAGACCTG  180
TTGGCACTGA GTGATACACT GGCAGTGATG TCTGTGGACC TGGGCAGTGA GTCCATGAAG  240
```

GTGGCCATTG TCAAACCTGG AGTGCCCATG GAAATTGTCT TGAATAAGGA ATCTCGGAGG 300

AAAACACCGG TGATCGTGAC CCTGAAAGAA AATGAAAGAT TCTTTGGAGA CAGTGCAGCA 360

AGCATGGCGA TTAAGAATCC AAAGGCTACG CTACGTTACT TCCAGCACCT CCTGGGGAAG 420

CAGGCAGATA ACCCCCATGT AGCTCTTTAC CAGGCCCGCT TCCCGGAGCA CGAGCTGACT 480

TTCGACCCAC AGAGGCAGAC TGTGCACTTT CAGATCAGCT CGCAGCTGCA GTTCTCACCT 540

GAGGAAGTGT TGGGCATGGT TCTCAATTAT TCTCGTTCTC TAGCTGAAGA TTTTGCAGAG 600

CAGCCCATCA AGGATGCAGT GATCACCGTG CCAGTCTTCT TCAACCAGGC CGAGCGCCGA 660

GCTGTGCTGC AGGCTGCTCG TATGGCTGGC CTCAAAGTGC TGCAGCTCAT CAATGACAAC 720

ACCGCCACTG CCCTCAGCTA TGGTGTCTTC CGCCGGAAAG ATATTAACAC CACTGCCCAG 780

AATATCATGT TCTATGACAT GGGCTCAGGC AGCACCGTAT GCACCATTGT GACCTACCAG 840

ATGGTGAAGA CTAAGGAAGC TGGGATGCAG CCACAGCTGC AGATCCGGGG AGTAGGATTT 900

GACCGTACCC TGGGGGGCCT GGAGATGGAG CTCCGGCTTC GAGAACGCCT GGCTGGGCTT 960

TTCAATGAGC AGCGCAAGGG TCAGAGAGCA AAGGATGTGC GGGAGAACCC GCGTGCCATG 1020

GCCAAGCTGC TGCGTGAGGC TAATCGGCTC AAAACCGTCC TCAGTGCCAA CGCTGACCAC 1080

ATGGCACAGA TTGAAGGCCT GATGGATGAT GTGGACTTCA AGGCAAAAGT GACTCGTGTG 1140

GAATTTGAGG AGTTGTGTGC AGACTTGTTT GAGCGGGTGC CTGGGCCTGT ACAGCAGGCC 1200

CTCCAGAGTG CCGAAATGAG TCTGGATGAG ATTGAGCAGG TGATCCTGGT GGGTGGGGCC 1260

ACTCGGGTCC CCAGAGTTCA GGAGGTGCTG CTGAAGGCCG TGGGCAAGGA GGAGCTGGGG 1320

AAGAACATCA ATGCAGATGA AGCAGCCGCC ATGGGGGCAG TGTACCAGGC AGCTGCGCTC 1380

AGCAAAGCCT TTAAAGTGAA GCCATTTGTC GTCCGAGATG CAGTGGTCTA CCCCATCCTG 1440

GTGGAGTTCA CGAGGGAGGT GGAGGAGGAG CCTGGGATTC ACAGCCTGAA GCACAATAAA 1500

CGGGTACTCT TCTCTCGGAT GGGGCCCTAC CCTCAACGCA AAGTCATCAC CTTTAACCGC 1560

TACAGCCATG ATTTCAACTT CCACATCAAC TACGGCGACC TGGGCTTCCT GGGGCCTGAA 1620

GATCTTCGGG TATTTGGCTC CCAGAATCTG ACCACAGTGA AGCTAAAAGG GGTGGGTGAC 1680

AGCTTCAAGA AGTATCCTGA CTACGAGTCC AAGGGCATCA AGGCTCACTT CAACCTGGAT 1740

GAGAGTGGCG TGCTCAGTCT AGACAGGGTG GAGTCTGTAT TTGAGACACT GGTAGAGGAC 1800

AGCGCAGAAG AGGAATCTAC TCTCACCAAA CTTGGCAACA CCATTTCCAG CCTGTTTGGA 1860

GGCGGTACCA CACCAGATGC CAAGGAGAAT GGTACTGATA CTGTCCAGGA GGAAGAGGAG 1920

AGCCCTGCAG AGGGGAGCAA GGACGAGCCT GGGGAGCAGG TGGAGCTCAA GGAGGAAGCT 1980

GAGGCCCCAG TGGAGGATGG CTCTCAGCCC CCACCCCCTG AACCTAAGGG AGATGCAACC 2040

CCTGAGGGAG AAAAGGCCAC AGAAAAAGAA AATGGGGACA AGTCTGAGGC CCAGAAACCA 2100

AGTGAGAAGG CAGAGGCAGG GCCTGAGGGC GTCGCTCCAG CCCCAGAGGG AGAGAAGAAG 2160

CAGAAGCCCG CCAGGAAGCG GCGAATGGTA GAGGAGATCG GGGTGGAGCT GGTTGTTCTG 2220

GACCTGCCTG ACTTGCCAGA GGATAAGCTG GCTCAGTCGG TGCAGAAACT TCAGGACTTG 2280

ACACTCCGAG ACCTGGAGAA GCAGGAACGG GAAAAAGCTG CCAACAGCTT GGAAGCGTTC 2340

ATATTTGAGA CCCAGGACAA GCTGTACCAG CCCGAGTACC AGGAAGTGTC CACAGAGGAG 2400

CAGCGTGAGG AGATCTCTGG GAAGCTCAGC GCCGCATCCA CCTGGCTGGA GGATGAGGGT 2460

GTTGGAGCCA CCACAGTGAT GTTGAAGGAG AAGCTGGCTG AGCTGAGGAA GCTGTGCCAA 2520

GGGCTGTTTT TTCGGGTAGA GGAGCGCAAG AAGTGGCCCG AACGGCTGTC TGCCCTCGAT 2580

AATCTCCTCA ACCATTCCAG CATGTTCCTC AAGGGGGCCC GGCTCATCCC AGAGATGGAC 2640

CAGATCTTCA CTGAGGTGGA GATGACAACG TTAGAGAAAG TCATCAATGA GACCTGGGCC 2700

TGGAAGAATG CAACTCTGGC CGAGCAGGCT AAGCTGCCCG CCACAGAGAA GCCTGTGTTG 2760

CTCTCAAAAG ACATTGAAGC TAAGATGATG GCCCTGGACC GAGAGGTGCA GTATCTGCTC 2820

AATAAGGCCA AGTTTACCAA GCCCCGGCCC CGGCCTAAGG ACAAGAATGG GACCCGGGCA 2880

GAGCCACCCC TCAATGCCAG TGCCAGTGAC CAGGGGGAGA AGGTCATCCC TCCAGCAGGC 2940

CAGACTGAAG ATGCAGAGCC CATTTCAGAA CCTGAGAAAG TAGAGACTGG ATCCGAGCCA 3000

GGAGACACTG AGCCTTTGGA GTTAGGAGGT CCTGGAGCAG AACCTGAACA GAAAGAACAA 3060

TCGACAGGAC AGAAGCGGCC TTTGAAGAAC GACGAACTAT AACCCCCACC TCTGTTTTCC 3120

CCATTCATCT CCACCCCCTT CCCCCACCAC TTCTATTTAT TTAACATCGA GGGTTGGGGG 3180

AGGGGTTGGT CCTGCCCTCG GCTGGAGTTC CTTTCTCACC CCTGTGATTT GGAGGTGTGG 3240

AGAAGGGGAA GGGAGGGACA GCTCACTGGT TCCTTCTGCA GTACCTCTGT GGTTAAAAAT 3300

GGAAACTGTT CTCCTCCCCA GCCCCACTCC CTGTTCCCTA CCCATATAGG CCCTAAATTT 3360

GGGAAAAATC ACTATTAATT TCTGAATCCT TTGCCTGTGG GTAGGAAGAG AATGGCTGCC 3420

AGTGGCTGAT GGGTCCCGGT GATGGGAAGG GTATCAGGTT GCTGGGGAGT TTCCACTCTT 3480

CTCTGGTGAT TGTTCCTTCC CTCCCTTCCT CTCCCACCAT GCGATGAGCA TCCTTTCAGG 3540

CCAGTGTCTG CAGAGCCTCA GTTACCAGGT TTGGTTTCTG AGTGCCTATC TGTGCTCTTT 3600

CCTCCCTCTG CGGGCTTCTC TTGCTCTGAG CCTCCCTTCC CCATTCCCAT GCAGCTCCTT 3660

TCCCCCTGGG TTTCCTTGGC TTCCTGCAGC AAATTGGGCA GTTCTCTGCC CCTTGCCTAA 3720

AAGCCTGTAC CTCTGGATTG GCGGAAGTAA ATCTGGAAGG ATTCTCACTC GTATTTCCCA 3780

CCCCTAGTGG CCAGAGGAGG GAGGGGCACA GTGAAGAAGG GAGCCCACCA CCTCTCCGAA 3840

GAGGAAAGCC ACGTAGAGTG GTTGGCATGG GGTGCCAGCA TCGTGCAAGC TCTGTCATAA 3900

TCTGCATCTT CCCAGCAGCC TGGTACCCCA GGTTCCTGTA ACTCCCTGCC TCCTCCTCTC 3960

TTCTGCTGTT CTGCTCCTCC CAGACAGAGC CTTTCCCTCA CCCCCTGACC CCCTGGGCTG 4020

ACCAAAATGT GCTTTCTACT GTGAGTCCCT ATCCCAAGAT CCTGGGGAAA GGAGAGACCA 4080

TGGTGTGAAT GTAGAGATGC CACCTCCCTC TCTCTGAGGC AGGCCTGTGG ATGAAGGAGG 4140

AGGGTCAGGG CTGGCCTTCC TCTGTGCATC ACTCTGCTAG GTTGGGGGCC CCCGACCCAC 4200

CATACCTACG CCTAGGGAGC CCGTCCTCCA GTATTCCGTC TGTAGCAGGA GCTAGGGCTG 4260

CTGCCTCAGC TCCAAGACAA GAATGAACCT GGCTGTTGCA GTCATTTTGT CTTTTCCTTT 4320

TTTTTTTTTT GCCACATTGG CAGAGATGGG ACCTAAGGGT CCCACCCCTC ACCCCACCCC 4380

CACCTCTTCT GTATGTTTGA ATTCTTTCAG TAGCTGTTGA TGCTGGTTGG ACAGGTTTGA 4440

GTCAAATTGT ACTTTGCTCC ATTGTTAATT GAGAAACTGT TTCAATAAAA TATTCTTTTC 4500

TAC 4503

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ala Ala Thr Val Arg Arg Gln Arg Pro Arg Arg Leu Leu Cys Trp
 5                  10                  15
Ala Leu Val Ala Val Leu Leu Ala Asp Leu Leu Ala Leu Ser Asp Thr
            20                  25                  30
Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
        35                  40                  45
Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu Ser
    50                  55                  60
Arg Arg Lys Thr Pro Val Thr Val Thr Leu Lys Glu Asn Glu Arg Phe
65                  70                  75                  80
Leu Gly Asp Ser Ala Ala Gly Met Ala Ile Lys Asn Pro Lys Ala Thr
                85                  90                  95
```

Leu Arg Tyr Phe Gln His Leu Leu Gly Lys Gln Ala Asp Asn Pro His
100 105 110
Val Ala Leu Tyr Arg Ser Arg Phe Pro Glu His Glu Leu Asn Val Asp
115 120 125
Pro Gln Arg Gln Thr Val Arg Phe Gln Ile Ser Pro Gln Leu Gln Phe
130 135 140
Ser Pro Glu Glu Val Leu Gly Met Val Leu Asn Tyr Ser Arg Ser Leu
145 150 155 160
Ala Glu Asp Phe Ala Glu Gln Pro Ile Lys Asp Ala Val Ile Thr Val
165 170 175
Pro Ala Phe Phe Asn Gln Ala Glu Arg Arg Ala Val Leu Gln Ala Ala
180 185 190
Arg Met Ala Gly Leu Lys Val Leu Gln Leu Ile Asn Asp Asn Thr Ala
195 200 205
Thr Ala Leu Ser Tyr Gly Val Phe Arg Arg Lys Asp Ile Asn Ser Thr
210 215 220
Ala Gln Asn Ile Met Phe Tyr Asp Met Gly Ser Gly Ser Thr Val Cys
225 230 235 240
Thr Ile Val Thr Tyr Gln Thr Val Lys Thr Lys Glu Ala Gly Thr Gln
245 250 255
Pro Gln Leu Gln Ile Arg Gly Val Gly Phe Asp Arg Thr Leu Gly Gly
260 265 270
Leu Glu Met Glu Leu Arg Leu Arg Glu His Leu Ala Lys Leu Phe Asn
275 280 285
Glu Gln Arg Lys Gly Gln Lys Ala Lys Asp Val Arg Glu Asn Pro Arg
290 295 300
Ala Met Ala Lys Leu Leu Arg Glu Ala Asn Arg Leu Lys Thr Val Leu
305 310 315 320
Ser Ala Asn Ala Asp His Met Ala Gln Ile Glu Gly Leu Met Asp Asp
325 330 335
Val Asp Phe Lys Ala Lys Val Thr Arg Val Glu Phe Glu Glu Leu Cys
340 345 350
Ala Asp Leu Phe Asp Arg Val Pro Gly Pro Val Gln Gln Ala Leu Gln
355 360 365
Ser Ala Glu Met Ser Leu Asp Gln Ile Glu Gln Val Ile Leu Val Gly
370 375 380
Gly Pro Thr Arg Val Pro Lys Val Gln Glu Val Leu Leu Lys Pro Val
385 390 395 400
Gly Lys Glu Glu Leu Gly Lys Asn Ile Asn Ala Asp Glu Ala Ala Ala
405 410 415
Met Gly Ala Val Tyr Gln Ala Ala Ala Leu Ser Lys Ala Phe Lys Val
420 425 430
Lys Pro Phe Val Val Arg Asp Ala Val Ile Tyr Pro Ile Leu Val Glu
435 440 445
Phe Thr Arg Glu Val Glu Glu Glu Pro Gly Leu Arg Ser Leu Lys His
450 455 460
Asn Lys Arg Val Leu Phe Ser Arg Met Gly Pro Tyr Pro Gln Arg Lys
465 470 475 480
Val Ile Thr Phe Asn Arg Tyr Ser His Asp Phe Asn Phe His Ile Asn
485 490 495
Tyr Gly Asp Leu Gly Phe Leu Gly Pro Glu Asp Leu Arg Val Phe Gly
500 505 510
Ser Gln Asn Leu Thr Thr Val Lys Leu Lys Gly Val Gly Glu Ser Phe
515 520 525
Lys Lys Tyr Pro Asp Tyr Glu Ser Lys Gly Ile Lys Ala His Phe Asn
530 535 540

Leu Asp Glu Ser Gly Val Leu Ser Leu Asp Arg Val Glu Ser Val Phe
545 550 555 560
Glu Thr Leu Val Glu Asp Ser Pro Glu Glu Glu Ser Thr Leu Thr Lys
565 570 575
Leu Gly Asn Thr Ile Ser Ser Leu Phe Gly Gly Gly Thr Ser Ser Asp
580 585 590
Ala Lys Glu Asn Gly Thr Asp Ala Val Gln Glu Glu Glu Glu Ser Pro
595 600 605
Ala Glu Gly Ser Lys Asp Glu Pro Ala Glu Gln Gly Glu Leu Lys Glu
610 615 620
Glu Ala Glu Ala Pro Met Glu Asp Thr Ser Gln Pro Pro Pro Ser Glu
625 630 635 640
Pro Lys Gly Asp Ala Ala Arg Glu Gly Glu Thr Pro Asp Glu Lys Glu
645 650 655
Ser Gly Asp Lys Ser Glu Ala Gln Lys Pro Asn Glu Lys Gly Gln Ala
660 665 670
Gly Pro Glu Gly Val Pro Pro Ala Pro Glu Glu Glu Lys Lys Gln Lys
675 680 685
Pro Ala Arg Lys Gln Lys Met Val Glu Glu Ile Gly Val Glu Leu Ala
690 695 700
Val Leu Asp Leu Pro Asp Leu Pro Glu Asp Glu Leu Ala His Ser Val
705 710 715 720
Gln Lys Leu Glu Asp Leu Thr Leu Arg Asp Leu Glu Lys Gln Glu Arg
725 730 735
Glu Lys Ala Ala Asn Ser Leu Glu Ala Phe Ile Phe Glu Thr Gln Asp
740 745 750
Lys Leu Tyr Gln Pro Glu Tyr Gln Glu Val Ser Thr Glu Glu Gln Arg
755 760 765
Glu Glu Ile Ser Gly Lys Leu Ser Ala Thr Ser Thr Trp Leu Glu Asp
770 775 780
Glu Gly Phe Gly Ala Thr Thr Val Met Leu Lys Asp Lys Leu Ala Glu
785 790 795 800
Leu Arg Lys Leu Cys Gln Gly Leu Phe Phe Arg Val Glu Glu Arg Arg
805 810 815
Lys Trp Pro Glu Arg Leu Ser Ala Leu Asp Asn Leu Leu Asn His Ser
820 825 830
Ser Ile Phe Leu Lys Gly Ala Arg Leu Ile Pro Glu Met Asp Gln Ile
835 840 845
Phe Thr Asp Val Glu Met Thr Thr Leu Glu Lys Val Ile Asn Asp Thr
850 855 860
Trp Thr Trp Lys Asn Ala Thr Leu Ala Glu Gln Ala Lys Leu Pro Ala
865 870 875 880
Thr Glu Lys Pro Val Leu Leu Ser Lys Asp Ile Glu Ala Lys Met Met
885 890 895
Ala Leu Asp Arg Glu Val Gln Tyr Leu Leu Asn Lys Ala Lys Phe Thr
900 905 910
Lys Pro Arg Pro Arg Pro Lys Asp Lys Asn Gly Thr Arg Thr Glu Pro
915 920 925
Pro Leu Asn Ala Ser Ala Gly Asp Gln Glu Glu Lys Val Ile Pro Pro
930 935 940
Thr Gly Gln Thr Glu Glu Ala Lys Ala Ile Leu Glu Pro Asp Lys Glu
945 950 955 960
Gly Leu Gly Thr Glu Ala Ala Asp Ser Glu Pro Leu Glu Leu Gly Gly
965 970 975
Pro Gly Ala Glu Ser Glu Gln Ala Glu Gln Thr Ala Gly Gln Lys Arg
980 985 990

Pro Leu Lys Asn Asp Glu Leu
995

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) IDENTIFICATION METHOD: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGAGGATGGA GCAGCGGTCG GGCCGCGGCT CCTAGGGGAG GCAGCGTGCT AGCTTCGGGG 60

GCGGGCCAGT AGCGGGAGCG AGGGCCGTAC GGACACCGGT CCCTTCGGCC TTGAAGTTCA 120

GGCGCTGAGC TGCCCCCTCG CGCTCGGGGT GGGCCGGAAT CCATTTCTGG GAGTGGGATC 180

TTCCACCTTC ATCAGGGTCA CAATGGCAGC TACAGTAAGG AGGCAGAGGC CAAGGAGGCT 240

ACTCTGTTGG GCCTTGGTGG CTGTCCTCTT GGCAGACCTG TTGGCACTGA GTGACACACT 300

GGCTGTGATG TCTGTGGACC TGGGCAGTGA ATCCATGAAG GTGGCCATTG TCAAGCCTGG 360

AGTGCCCATG GAGATTGTAT TGAACAAGGA ATCTCGGAGG AAAACTCCGG TGACTGTGAC 420

CTTGAAGGAA AACGAAAGGT TCTAGGTGA CAGTGCAGCT GGCATGGCCA TCAAGAACCC 480

AAAGGCTACG CTCCGTTATT TCCAGCACCT CCTTGGAAAG CAGGCAGATA ACCCTCATGT 540

GGCTCTTTAC CGGTCCCGTT TCCCAGAACA TGAGCTCAAT GTTGACCCAC AGAGGCAGAC 600

TGTGCGCTTC CAGATCAGTC CGCAGCTGCA GTTCTCTCCC GAGGAGGTGC TGGGCATGGT 660

TCTCAACTAC TCCCGTTCCC TGGCTGAAGA TTTTGCAGAA CAACCTATTA AGGATGCAGT 720

GATCACCGTG CCAGCCTTTT TCAACCAGGC CGAGCGCCGA GCTGTGCTGC AGGCTGCTCG 780

TATGGCTGGC CTCAAGGTGC TGCAGCTCAT CAATGACAAC ACTGCCACAG CCCTCAGCTA 840

TGGTGTCTTC CGCCGGAAAG ATATCAATTC CACTGCACAG AATATCATGT TCTATGACAT 900

GGGCTCGGGC AGCACTGTGT GTACCATCGT GACCTACCAA ACGGTGAAGA CTAAGGAGGC 960

TGGGACGCAG CCACAGCTAC AGATCCGGGG CGTGGGATTT GACCGCACCC TGGGTGGCCT 1020

GGAGATGGAG CTTCGGCTGC GAGAGCACCT GGCTAAGCTC TTCAATGAGC AGCGCAAGGG 1080

CCAGAAAGCC AAGGATGTTC GGGAAAACCC CCGAGCCATG GCCAAACTGC TTCGGGAAGC 1140

CAATCGGCTT AAAACCGTCC TGAGTGCCAA TGCTGATCAC ATGGCACAGA TTGAAGGCTT 1200

GATGGACGAT GTGGACTTCA AGGCAAAAGT AACTCGAGTG GAGTTTGAGG AGCTGTGTGC 1260

AGATTTGTTT GATCGAGTGC CTGGGCCTGT ACAGCAGGCC CTGCAGAGTG CTGAGATGAG 1320

CCTGGATCAA ATTGAGCAGG TGATCCTGGT GGGTGGGCCC ACTCGTGTTC CCAAAGTTCA 1380

AGAGGTGCTG CTGAAGCCTG TGGGCAAGGA GGAACTAGGA AAGAACATCA ATGCCGATGA 1440

AGCAGCTGCC ATGGGGGCCG TGTACCAGGC AGCGGCACTG AGCAAAGCCT TCAAAGTGAA 1500

GCCATTTGTT GTGCGTGATG CTGTTATTTA CCCCATCCTG GTGGAGTTCA CAAGGGAGGT 1560

GGAGGAGGAG CCTGGGCTTC GAAGCCTGAA GCACAATAAA CGTGTGCTCT TCTCCCGAAT 1620

GGGGCCCTAC CCTCAGCGCA AAGTCATCAC CTTTAACCGA TACAGCCATG ATTTCAACTT 1680

TCACATCAAC TACGGTGACC TGGGCTTCCT GGGGCCTGAG GATCTTCGGG TATTTGGCTC 1740

CCAGAATCTG ACCACAGTGA AACTAAAAGG TGTGGGAGAG AGCTTCAAGA AATATCCTGA 1800

CTATGAGTCC AAAGGCATCA AGGCCCACTT TAACCTAGAC GAGAGTGGAG TGCTCAGTTT 1860

AGACAGGGTG GAGTCCGTAT TCGAGACCCT GGTGGAGGAC AGCCCAGAGG AAGAGTCTAC 1920

TCTTACCAAA CTTGGCAACA CCATTTCCAG CCTGTTTGGC GGTGGTACCT CATCAGATGC 1980

CAAAGAGAAT GGTACTGATG CTGTACAGGA GGAGGAGGAG AGCCCTGCTG AGGGGAGCAA 2040

GGATGAGCCT GCAGAACAGG GGGAACTCAA GGAGGAAGCT GAAGCCCCAA TGGAGGATAC 2100

CTCCCAGCCT CCACCCTCTG AGCCTAAGGG GGATGCAGCC CGTGAGGGAG AAACACCTGA 2160

TGAAAAAGAA AGTGGGGACA AGTCTGAGGC CCAGAAGCCC AATGAGAAGG GGCAGGCAGG 2220

GCCTGAGGGT GTCCCTCCAG CTCCCGAGGA AGAAAAAAAG CAGAAACCTG CCCGGAAGCA 2280

GAAAATGGTG GAGGAGATAG GTGTGGAACT GGCTGTCTTG GACCTGCCAG ACTTGCCAGA 2340

GGATGAGCTG GCCCATTCCG TGCAGAAACT TGAGGACTTG ACCCTGCGAG ACCTTGAAAA 2400

GCAGGAGAGG GAGAAAGCTG CCAACAGCTT AGAAGCTTTT ATCTTTGAGA CCCAGGACAA 2460

ACTGTACCAA CCTGAGTACC AGGAAGTGTC CACTGAGGAA CAACGGGAGG AGATCTCTGG 2520

AAAACTCAGT GCCACTTCTA CCTGGCTGGA GGATGAGGGA TTTGGAGCCA CCACTGTGAT 2580

GTTGAAGGAC AAGCTGGCTG AGCTGAGAAA GCTGTGCCAA GGGCTGTTTT TTCGGGTGGA 2640

AGAGCGCAGG AAATGGCCAG AGCGGCTTTC AGCTCTGGAT AATCTCCTCA ATCACTCCAG 2700

CATTTTCCTC AAGGGTGCCC GACTCATCCC AGAGATGGAC CAGATCTTCA CTGACGTGGA 2760

GATGACAACG TTGGAGAAAG TCATCAATGA CACCTGGACC TGGAAGAATG CAACCCTGGC 2820

CGAGCAGGCC AAGCTTCCTG CCACAGAGAA ACCCGTGCTG CTTTCAAAAG ACATCGAGGC 2880

CAAAATGATG GCCCTGGACC GGGAGGTGCA GTATCTACTC AATAAGGCCA AGTTTACTAA 2940

ACCCCGGCCA CGGCCCAAGG ACAAGAATGG CACCCGGACA GAGCCTCCCC TCAATGCCAG 3000

TGCTGGTGAC CAAGAGGAAA AGGTCATTCC ACCTACAGGC CAGACTGAAG AGGCGAAGGC 3060

CATCTTAGAA CCTGACAAAG AAGGGCTTGG TACAGAGGCA GCAGACTCTG AGCCTCTGGA 3120

ATTAGGAGGT CCTGGTGCAG AATCTGAACA GGCAGAGCAG ACAGCAGGGC AGAAGCGGCC 3180

TTTGAAGAAT GATGAGCTGT GACCCCGCGC CTCCGCTCCA CTTGCCTCCA GCCCCTTCTC 3240

CTACCACCTC TA 3252

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
5 10 15
Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATACGACTC ACTATAGGGA 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Lys Pro Gly Val Pro Met Glu
5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AARCCiGGiG TNCCNATGGA 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu
5 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCACCCTTGA GGAAAATGCT 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCCAGAAGCC CAATGAGAAG 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2861 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAAAGAAGTA GACATGGGAG ACTTCATTTT GTTCTGTACT AAGAAAAATT CTTCTGCCTT 60

GGGATGCTGT TGATCTATGA CCTTACCCCC AACCCTGTGC TCTCTGAAAC ATGTGCTGTG 120

TCCACTCAGG GTTAAATGGA TTAAGGGCGG TGCAAGATGT GCTTTGTTAA ACAGATGCTT 180

GAAGGCAGCA TGCTCGTTAG GAGTCATCAC CACTCCCTAA TCTCAAGTAC CCAGGGACAC 240

AAACACTGCG GAAGGCCACA GGGTCCTCTG CCTAGGAAAG CCAGAGACCT TTGTTCACTT 300

GTTTATCTGC TGACCTTCCC TCCACTATTG TCCTATGACC CTGCCAAATC CCCCTCTGCC 360

AGAAACACCC AAGAATGATC AATAAAAAAA AAAAAAAAAA AAAAAGGAAG AATAGACTCT 420

CTCTGGGACT GCCAATAATT TTTCCTTCTA AGCATAGACA CCGGACCACT CTCCACCTAA 480

GCATCACGAA AAATGTAGAG AAAGGAAGAG CTAAGAGCTC CTTAAACAAG TTCAGGCTTG 540

ACACAACCCT GGCCCTGACA GCCAGGGTCT TCAAGCGGGC CTTTCTGTGA AGGGTGGCCA 600

GGCATCAACT TAGTAGGAGA GAAAACAGAT GACTTATTTC CATCCACACT TAAGGAAAAT 660

GCAGTCTCCA AGGACTGCGT ACATTTCTTT TTCGAGAAGG AGTCTCGCTG TTGTCGCCCA 720

GGCTGGAGTG CAGTGGCGCA GTCTGGGCTC ACAGCAACCT CTGCCTCCCG GATTCAAGCA 780

ATTCTCCTGC CTCAGCCTCG TGAGTAGCTG GGATTACAGG CACCCGCCAC CACGCCTGGC 840

TAATTTTTGT AGTTTTGGTA GAGACGGGGT TTCACCATGT TGGCCAGGCT GGTCTCGAAC 900

TCCTGACCTC CAGTGATTCG CCCGCCTTGG CCTCCCAAAA TGCTGGGATT ACAGGCGTGA 960

GCCACCGCGC CCGGGCGACT GCGCACATTT CTATGGAGCT GTAAGTTAAA AGAGAAGGCA 1020

GTGAGGTGCT TCTGTCATTC TATGACAGAA ACAGCTAAAG AGTAGAGAAA TGTTCACAAG 1080

ATTTAATAGA ACAGAAATAG GAGAAGGTGC ACACAAGCTC AACCAACTAT AGCCTCACAA 1140

ATAAAAGTGT CTTTTGTGTG TAGTACTTAA GTTTGGAATA TTCTTTCTTA TACAAATGAG 1200

TGGGGCTTAA CCTAAGAAAT CCTGGCCAGA TTCTGCGACG AATGCATCGG TTATCTCTGA 1260

CCCATCAGCA AACATCTTTT TCTGTGGCTT CAGTTTCCTC AGTAAAACAG AGGGGGTTGC 1320

GACGGACTCA GTCCGAGGCA CAGCCATTCT CCAACGTCTA TCCAAAGCCT AGGGCACCTC 1380

AATACTAACC GGCAGGCCAG CGCCCCCTCC GCGGGGCTGC GGACAGGACG CCTGTTATTC 1440

CATTCCTCGG CCGGGCTCTA CAGGTGACCG GAAGAAGAGC CCCGAGTGCG GGACTGCAGT 1500

GCGCCCGACC TGCTCTAGGC GCAGGTCACT CCCGAACCCC GGCAGCAAAG CATCCAGCGC 1560

CGGAAAAGGT CCCGCGGTCG CCCCGGGGCC GGCGCTGGGG AGGAAGGAGT GGAGCGCGCT 1620

GGCCCCGTGA CGTGGTCCAA TCCCAGGCCG ACGCCGGCTG CTTCTGCCCA ACCGGTGGCT 1680

GGTCCCCTCC GCCGCCCCCA TTACAAGGCT GGCAAAGGGA GGGGGCGGGG CCTGGGACGT 1740

GGTCCAATGA GTACGCGCGC CGGGGCGGCG GGGGCGGGGC CGGGCGCGCA GCGCAGGGCC 1800

GGGCGGCCGA GGCTCCAATG AGCGCCCGCC GCGTCCGGGG CCGGCTGGTG CGCGAGACGC 1860

CGCCGAGAGG TTGGTGGCTA ATGTAACAGT TTGCAAACCG AGAGGAGTTG TGAAGGGCGC 1920

GGGTGGGGGG CGCTGCCGGC CTCGTGGGTA CGTTCGTGCC GCGTCTGTCC CAGAGCTGGG 1980

GCCGCAGGAG CGGAGGCAAG AGGTAGCGGG GGTGGATGGA GGTGCGGGCC GGCCACCCCT 2040

CCTAGGGGAG ACAGCGTGCG AGCTCCGGGG GCGGGTCGGG AGCGCAAGGG AGGGCCGCGC 2100

GGACGCCGGG CGCTCGGCCT CGCACCGGGG GGCACGCAGC TCGGCCCCCG GTCTGTCCCC 2160

ACTTGCTGGG GCGGGCCGGG ATCCGTTTCC GGGAGTGGGA GCCGCCGCCT TCGTCAGGTG 2220

GGGTTTAGGT GAACACCGGG TAACGGCTAC CCGCCGGGCG GGGAACCTTA CCGCCCCTGG 2280

CACTGCGTCT GTGGGCACAG CGGGGCCGGG GAGTGAGCTG GGAAAGGGGA GGGGGCGGGA 2340

CAACCCGCAG GGATGCCGAG GAGGAGATAG GCCTTTCCTT CATCCTAGCT ACCCCCAACG 2400

TCATTACCTT TCTCTTCCCG TCCAGGCCCA GCTGGCTTTC CCCGTCAGCG GGGGAGCTCC 2460

AGGTGTGGGG AGGTGGTTGA GCCCTGGGCG GGGATCCCTG GCCGCACCCC AGGTGTCTGA 2520

```
CAACAGGCAC AGTGCTGCGG TGCGCCACTC ACTGCCTGTG TGGTGGACAA AAGGCTCGGG 2580
TCTCCTTTCT CTTGTCCTGT TAGCTTCTCT GTTTAGGGAT GTGGCAAAGC CGAGGACCCA 2640
TGCTCTTTCA CTTGGGCCTT TGTGTGGGCG CTGCTGGGAT GATTAGAGAA TGGTTTGTAC 2700
CCATCAGGAG GGAGAAGGGG AGAAGTAGGC TGATCTGCCC TGGGTAAGAA TGAAGTAGAT 2760
ATGAATCTTA CAGCCTCTCC GTTCTGGGAT GTGATTCTGT CTCCTTCACT CCGGGTATCC 2820
AGTTTTAAGT GTTTTCTTTC TTCGCCTCCC CCAGGGGCAC T                      2861
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: HSP Research Institute, Inc.
(B) STREET: 2-8, Doshomachi 2-chome, Chuo-ku,
(C) CITY: Osaka-shi, Osaka
(E) COUNTRY: JP
(F) POSTAL CODE (ZIP): none

(ii) TITLE OF INVENTION: STRESS PROTEINS

(iii) NUMBER OF SEQUENCES: 12

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: EP 96 12 0622.0
(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 7-349661
(B) FILING DATE: 20-DEC-1995

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 8-213181
(B) FILING DATE: 23-JUL-1996

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Met Ala Asp Lys Val Arg Arg Gln Arg Pro Arg Arg Arg Val Cys Trp
 1               5                  10                  15

Ala Leu Val Ala Val Leu Leu Ala Asp Leu Leu Ala Leu Ser Asp Thr
            20                  25                  30

Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
        35                  40                  45

Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu Ser
    50                  55                  60

Arg Arg Lys Thr Pro Val Ile Val Thr Leu Lys Glu Asn Glu Arg Phe
65                  70                  75                  80

Phe Gly Asp Ser Ala Ala Ser Met Ala Ile Lys Asn Pro Lys Ala Thr
                85                  90                  95
```

Leu Arg Tyr Phe Gln His Leu Leu Gly Lys Gln Ala Asp Asn Pro His
100 105 110

Val Ala Leu Tyr Gln Ala Arg Phe Pro Glu His Glu Leu Thr Phe Asp
115 120 125

Pro Gln Arg Gln Thr Val His Phe Gln Ile Ser Ser Gln Leu Gln Phe
130 135 140

Ser Pro Glu Glu Val Leu Gly Met Val Leu Asn Tyr Ser Arg Ser Leu
145 150 155 160

Ala Glu Asp Phe Ala Glu Gln Pro Ile Lys Asp Ala Val Ile Thr Val
165 170 175

Pro Val Phe Phe Asn Gln Ala Glu Arg Arg Ala Val Leu Gln Ala Ala
180 185 190

Arg Met Ala Gly Leu Lys Val Leu Gln Leu Ile Asn Asp Asn Thr Ala
195 200 205

Thr Ala Leu Ser Tyr Gly Val Phe Arg Arg Lys Asp Ile Asn Thr Thr
210 215 220

Ala Gln Asn Ile Met Phe Tyr Asp Met Gly Ser Gly Ser Thr Val Cys
225 230 235 240

Thr Ile Val Thr Tyr Gln Met Val Lys Thr Lys Glu Ala Gly Met Gln
245 250 255

Pro Gln Leu Gln Ile Arg Gly Val Gly Phe Asp Arg Thr Leu Gly Gly
260 265 270

Leu Glu Met Glu Leu Arg Leu Arg Glu Arg Leu Ala Gly Leu Phe Asn
275 280 285

Glu Gln Arg Lys Gly Gln Arg Ala Lys Asp Val Arg Glu Asn Pro Arg
290 295 300

Ala Met Ala Lys Leu Leu Arg Glu Ala Asn Arg Leu Lys Thr Val Leu
305 310 315 320

Ser Ala Asn Ala Asp His Met Ala Gln Ile Glu Gly Leu Met Asp Asp
325 330 335

Val Asp Phe Lys Ala Lys Val Thr Arg Val Glu Phe Glu Glu Leu Cys
340 345 350

Ala Asp Leu Phe Glu Arg Val Pro Gly Pro Val Gln Gln Ala Leu Gln
355 360 365

Ser Ala Glu Met Ser Leu Asp Glu Ile Glu Gln Val Ile Leu Val Gly
370 375 380

Gly Ala Thr Arg Val Pro Arg Val Gln Glu Val Leu Leu Lys Ala Val
385 390 395 400

Gly Lys Glu Glu Leu Gly Lys Asn Ile Asn Ala Asp Glu Ala Ala Ala
405 410 415

```
Met Gly Ala Val Tyr Gln Ala Ala Ala Leu Ser Lys Ala Phe Lys Val
            420                 425                 430

Lys Pro Phe Val Val Arg Asp Ala Val Val Tyr Pro Ile Leu Val Glu
        435                 440                 445

Phe Thr Arg Glu Val Glu Glu Glu Pro Gly Ile His Ser Leu Lys His
    450                 455                 460

Asn Lys Arg Val Leu Phe Ser Arg Met Gly Pro Tyr Pro Gln Arg Lys
465                 470                 475                 480

Val Ile Thr Phe Asn Arg Tyr Ser His Asp Phe Asn Phe His Ile Asn
                485                 490                 495

Tyr Gly Asp Leu Gly Phe Leu Gly Pro Glu Asp Leu Arg Val Phe Gly
            500                 505                 510

Ser Gln Asn Leu Thr Thr Val Lys Leu Lys Gly Val Gly Asp Ser Phe
        515                 520                 525

Lys Lys Tyr Pro Asp Tyr Glu Ser Lys Gly Ile Lys Ala His Phe Asn
    530                 535                 540

Leu Asp Glu Ser Gly Val Leu Ser Leu Asp Arg Val Glu Ser Val Phe
545                 550                 555                 560

Glu Thr Leu Val Glu Asp Ser Ala Glu Glu Glu Ser Thr Leu Thr Lys
                565                 570                 575

Leu Gly Asn Thr Ile Ser Ser Leu Phe Gly Gly Gly Thr Thr Pro Asp
            580                 585                 590

Ala Lys Glu Asn Gly Thr Asp Thr Val Gln Glu Glu Glu Glu Ser Pro
        595                 600                 605

Ala Glu Gly Ser Lys Asp Glu Pro Gly Glu Gln Val Glu Leu Lys Glu
    610                 615                 620

Glu Ala Glu Ala Pro Val Glu Asp Gly Ser Gln Pro Pro Pro Pro Glu
625                 630                 635                 640

Pro Lys Gly Asp Ala Thr Pro Glu Gly Glu Lys Ala Thr Glu Lys Glu
                645                 650                 655

Asn Gly Asp Lys Ser Glu Ala Gln Lys Pro Ser Glu Lys Ala Glu Ala
            660                 665                 670

Gly Pro Glu Gly Val Ala Pro Ala Pro Glu Gly Glu Lys Lys Gln Lys
        675                 680                 685

Pro Ala Arg Lys Arg Arg Met Val Glu Glu Ile Gly Val Glu Leu Val
    690                 695                 700

Val Leu Asp Leu Pro Asp Leu Pro Glu Asp Lys Leu Ala Gln Ser Val
705                 710                 715                 720

Gln Lys Leu Gln Asp Leu Thr Leu Arg Asp Leu Glu Lys Gln Glu Arg
                725                 730                 735
```

Glu Lys Ala Ala Asn Ser Leu Glu Ala Phe Ile Phe Glu Thr Gln Asp
740 745 750

Lys Leu Tyr Gln Pro Glu Tyr Gln Glu Val Ser Thr Glu Glu Gln Arg
755 760 765

Glu Glu Ile Ser Gly Lys Leu Ser Ala Ala Ser Thr Trp Leu Glu Asp
770 775 780

Glu Gly Val Gly Ala Thr Thr Val Met Leu Lys Glu Lys Leu Ala Glu
785 790 795 800

Leu Arg Lys Leu Cys Gln Gly Leu Phe Phe Arg Val Glu Glu Arg Lys
805 810 815

Lys Trp Pro Glu Arg Leu Ser Ala Leu Asp Asn Leu Leu Asn His Ser
820 825 830

Ser Met Phe Leu Lys Gly Ala Arg Leu Ile Pro Glu Met Asp Gln Ile
835 840 845

Phe Thr Glu Val Glu Met Thr Thr Leu Glu Lys Val Ile Asn Glu Thr
850 855 860

Trp Ala Trp Lys Asn Ala Thr Leu Ala Glu Gln Ala Lys Leu Pro Ala
865 870 875 880

Thr Glu Lys Pro Val Leu Leu Ser Lys Asp Ile Glu Ala Lys Met Met
885 890 895

Ala Leu Asp Arg Glu Val Gln Tyr Leu Leu Asn Lys Ala Lys Phe Thr
900 905 910

Lys Pro Arg Pro Arg Pro Lys Asp Lys Asn Gly Thr Arg Ala Glu Pro
915 920 925

Pro Leu Asn Ala Ser Ala Ser Asp Gln Gly Glu Lys Val Ile Pro Pro
930 935 940

Ala Gly Gln Thr Glu Asp Ala Glu Pro Ile Ser Glu Pro Glu Lys Val
945 950 955 960

Glu Thr Gly Ser Glu Pro Gly Asp Thr Glu Pro Leu Glu Leu Gly Gly
965 970 975

Pro Gly Ala Glu Pro Glu Gln Lys Glu Gln Ser Thr Gly Gln Lys Arg
980 985 990

Pro Leu Lys Asn Asp Glu Leu
995

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:103..3099

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
TTGTGAAGGG CGCGGGTGGG GGGCGCTGCC GGCCTCGTGG GTACGTTCGT GCCGCGTCTG      60

TCCCAGAGCT GGGGCCGCAG GAGCGGAGGC AAGAGGGGCA CT ATG GCA GAC AAA        114
                                                  Met Ala Asp Lys
                                                    1

GTT AGG AGG CAG AGG CCG AGG AGG CGA GTC TGT TGG GCC TTG GTG GCT        162
Val Arg Arg Gln Arg Pro Arg Arg Arg Val Cys Trp Ala Leu Val Ala
  5                  10                  15                  20

GTG CTC TTG GCA GAC CTG TTG GCA CTG AGT GAT ACA CTG GCA GTG ATG        210
Val Leu Leu Ala Asp Leu Leu Ala Leu Ser Asp Thr Leu Ala Val Met
                 25                  30                  35

TCT GTG GAC CTG GGC AGT GAG TCC ATG AAG GTG GCC ATT GTC AAA CCT        258
Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala Ile Val Lys Pro
             40                  45                  50

GGA GTG CCC ATG GAA ATT GTC TTG AAT AAG GAA TCT CGG AGG AAA ACA        306
Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu Ser Arg Arg Lys Thr
         55                  60                  65

CCG GTG ATC GTG ACC CTG AAA GAA AAT GAA AGA TTC TTT GGA GAC AGT        354
Pro Val Ile Val Thr Leu Lys Glu Asn Glu Arg Phe Phe Gly Asp Ser
     70                  75                  80

GCA GCA AGC ATG GCG ATT AAG AAT CCA AAG GCT ACG CTA CGT TAC TTC        402
Ala Ala Ser Met Ala Ile Lys Asn Pro Lys Ala Thr Leu Arg Tyr Phe
 85                  90                  95                 100

CAG CAC CTC CTG GGG AAG CAG GCA GAT AAC CCC CAT GTA GCT CTT TAC        450
Gln His Leu Leu Gly Lys Gln Ala Asp Asn Pro His Val Ala Leu Tyr
                105                 110                 115

CAG GCC CGC TTC CCG GAG CAC GAG CTG ACT TTC GAC CCA CAG AGG CAG        498
Gln Ala Arg Phe Pro Glu His Glu Leu Thr Phe Asp Pro Gln Arg Gln
            120                 125                 130

ACT GTG CAC TTT CAG ATC AGC TCG CAG CTG CAG TTC TCA CCT GAG GAA        546
Thr Val His Phe Gln Ile Ser Ser Gln Leu Gln Phe Ser Pro Glu Glu
        135                 140                 145

GTG TTG GGC ATG GTT CTC AAT TAT TCT CGT TCT CTA GCT GAA GAT TTT        594
Val Leu Gly Met Val Leu Asn Tyr Ser Arg Ser Leu Ala Glu Asp Phe
    150                 155                 160

GCA GAG CAG CCC ATC AAG GAT GCA GTG ATC ACC GTG CCA GTC TTC TTC        642
Ala Glu Gln Pro Ile Lys Asp Ala Val Ile Thr Val Pro Val Phe Phe
165                 170                 175                 180
```

```
AAC CAG GCC GAG CGC CGA GCT GTG CTG CAG GCT GCT CGT ATG GCT GGC          690
Asn Gln Ala Glu Arg Arg Ala Val Leu Gln Ala Ala Arg Met Ala Gly
                185                 190                 195

CTC AAA GTG CTG CAG CTC ATC AAT GAC AAC ACC GCC ACT GCC CTC AGC          738
Leu Lys Val Leu Gln Leu Ile Asn Asp Asn Thr Ala Thr Ala Leu Ser
            200                 205                 210

TAT GGT GTC TTC CGC CGG AAA GAT ATT AAC ACC ACT GCC CAG AAT ATC          786
Tyr Gly Val Phe Arg Arg Lys Asp Ile Asn Thr Thr Ala Gln Asn Ile
        215                 220                 225

ATG TTC TAT GAC ATG GGC TCA GGC AGC ACC GTA TGC ACC ATT GTG ACC          834
Met Phe Tyr Asp Met Gly Ser Gly Ser Thr Val Cys Thr Ile Val Thr
    230                 235                 240

TAC CAG ATG GTG AAG ACT AAG GAA GCT GGG ATG CAG CCA CAG CTG CAG          882
Tyr Gln Met Val Lys Thr Lys Glu Ala Gly Met Gln Pro Gln Leu Gln
245                 250                 255                 260

ATC CGG GGA GTA GGA TTT GAC CGT ACC CTG GGG GGC CTG GAG ATG GAG          930
Ile Arg Gly Val Gly Phe Asp Arg Thr Leu Gly Gly Leu Glu Met Glu
                265                 270                 275

CTC CGG CTT CGA GAA CGC CTG GCT GGG CTT TTC AAT GAG CAG CGC AAG          978
Leu Arg Leu Arg Glu Arg Leu Ala Gly Leu Phe Asn Glu Gln Arg Lys
            280                 285                 290

GGT CAG AGA GCA AAG GAT GTG CGG GAG AAC CCG CGT GCC ATG GCC AAG         1026
Gly Gln Arg Ala Lys Asp Val Arg Glu Asn Pro Arg Ala Met Ala Lys
        295                 300                 305

CTG CTG CGT GAG GCT AAT CGG CTC AAA ACC GTC CTC AGT GCC AAC GCT         1074
Leu Leu Arg Glu Ala Asn Arg Leu Lys Thr Val Leu Ser Ala Asn Ala
    310                 315                 320

GAC CAC ATG GCA CAG ATT GAA GGC CTG ATG GAT GAT GTG GAC TTC AAG         1122
Asp His Met Ala Gln Ile Glu Gly Leu Met Asp Asp Val Asp Phe Lys
325                 330                 335                 340

GCA AAA GTG ACT CGT GTG GAA TTT GAG GAG TTG TGT GCA GAC TTG TTT         1170
Ala Lys Val Thr Arg Val Glu Phe Glu Glu Leu Cys Ala Asp Leu Phe
                345                 350                 355

GAG CGG GTG CCT GGG CCT GTA CAG CAG GCC CTC CAG AGT GCC GAA ATG         1218
Glu Arg Val Pro Gly Pro Val Gln Gln Ala Leu Gln Ser Ala Glu Met
            360                 365                 370

AGT CTG GAT GAG ATT GAG CAG GTG ATC CTG GTG GGT GGG GCC ACT CGG         1266
Ser Leu Asp Glu Ile Glu Gln Val Ile Leu Val Gly Gly Ala Thr Arg
        375                 380                 385

GTC CCC AGA GTT CAG GAG GTG CTG CTG AAG GCC GTG GGC AAG GAG GAG         1314
Val Pro Arg Val Gln Glu Val Leu Leu Lys Ala Val Gly Lys Glu Glu
    390                 395                 400

CTG GGG AAG AAC ATC AAT GCA GAT GAA GCA GCC GCC ATG GGG GCA GTG         1362
Leu Gly Lys Asn Ile Asn Ala Asp Glu Ala Ala Ala Met Gly Ala Val
405                 410                 415                 420
```

```
TAC CAG GCA GCT GCG CTC AGC AAA GCC TTT AAA GTG AAG CCA TTT GTC      1410
Tyr Gln Ala Ala Ala Leu Ser Lys Ala Phe Lys Val Lys Pro Phe Val
                425                 430                 435

GTC CGA GAT GCA GTG GTC TAC CCC ATC CTG GTG GAG TTC ACG AGG GAG      1458
Val Arg Asp Ala Val Val Tyr Pro Ile Leu Val Glu Phe Thr Arg Glu
            440                 445                 450

GTG GAG GAG GAG CCT GGG ATT CAC AGC CTG AAG CAC AAT AAA CGG GTA      1506
Val Glu Glu Glu Pro Gly Ile His Ser Leu Lys His Asn Lys Arg Val
        455                 460                 465

CTC TTC TCT CGG ATG GGG CCC TAC CCT CAA CGC AAA GTC ATC ACC TTT      1554
Leu Phe Ser Arg Met Gly Pro Tyr Pro Gln Arg Lys Val Ile Thr Phe
    470                 475                 480

AAC CGC TAC AGC CAT GAT TTC AAC TTC CAC ATC AAC TAC GGC GAC CTG      1602
Asn Arg Tyr Ser His Asp Phe Asn Phe His Ile Asn Tyr Gly Asp Leu
485                 490                 495                 500

GGC TTC CTG GGG CCT GAA GAT CTT CGG GTA TTT GGC TCC CAG AAT CTG      1650
Gly Phe Leu Gly Pro Glu Asp Leu Arg Val Phe Gly Ser Gln Asn Leu
                505                 510                 515

ACC ACA GTG AAG CTA AAA GGG GTG GGT GAC AGC TTC AAG AAG TAT CCT      1698
Thr Thr Val Lys Leu Lys Gly Val Gly Asp Ser Phe Lys Lys Tyr Pro
            520                 525                 530

GAC TAC GAG TCC AAG GGC ATC AAG GCT CAC TTC AAC CTG GAT GAG AGT      1746
Asp Tyr Glu Ser Lys Gly Ile Lys Ala His Phe Asn Leu Asp Glu Ser
        535                 540                 545

GGC GTG CTC AGT CTA GAC AGG GTG GAG TCT GTA TTT GAG ACA CTG GTA      1794
Gly Val Leu Ser Leu Asp Arg Val Glu Ser Val Phe Glu Thr Leu Val
    550                 555                 560

GAG GAC AGC GCA GAA GAG GAA TCT ACT CTC ACC AAA CTT GGC AAC ACC      1842
Glu Asp Ser Ala Glu Glu Glu Ser Thr Leu Thr Lys Leu Gly Asn Thr
565                 570                 575                 580

ATT TCC AGC CTG TTT GGA GGC GGT ACC ACA CCA GAT GCC AAG GAG AAT      1890
Ile Ser Ser Leu Phe Gly Gly Gly Thr Thr Pro Asp Ala Lys Glu Asn
                585                 590                 595

GGT ACT GAT ACT GTC CAG GAG GAA GAG GAG AGC CCT GCA GAG GGG AGC      1938
Gly Thr Asp Thr Val Gln Glu Glu Glu Glu Ser Pro Ala Glu Gly Ser
            600                 605                 610

AAG GAC GAG CCT GGG GAG CAG GTG GAG CTC AAG GAG GAA GCT GAG GCC      1986
Lys Asp Glu Pro Gly Glu Gln Val Glu Leu Lys Glu Glu Ala Glu Ala
        615                 620                 625

CCA GTG GAG GAT GGC TCT CAG CCC CCA CCC CCT GAA CCT AAG GGA GAT      2034
Pro Val Glu Asp Gly Ser Gln Pro Pro Pro Pro Glu Pro Lys Gly Asp
    630                 635                 640

GCA ACC CCT GAG GGA GAA AAG GCC ACA GAA AAA GAA AAT GGG GAC AAG      2082
Ala Thr Pro Glu Gly Glu Lys Ala Thr Glu Lys Glu Asn Gly Asp Lys
645                 650                 655                 660
```

```
TCT GAG GCC CAG AAA CCA AGT GAG AAG GCA GAG GCA GGG CCT GAG GGC      2130
Ser Glu Ala Gln Lys Pro Ser Glu Lys Ala Glu Ala Gly Pro Glu Gly
                665                 670                 675

GTC GCT CCA GCC CCA GAG GGA GAG AAG AAG CAG AAG CCC GCC AGG AAG      2178
Val Ala Pro Ala Pro Glu Gly Glu Lys Lys Gln Lys Pro Ala Arg Lys
            680                 685                 690

CGG CGA ATG GTA GAG GAG ATC GGG GTG GAG CTG GTT GTT CTG GAC CTG      2226
Arg Arg Met Val Glu Glu Ile Gly Val Glu Leu Val Val Leu Asp Leu
        695                 700                 705

CCT GAC TTG CCA GAG GAT AAG CTG GCT CAG TCG GTG CAG AAA CTT CAG      2274
Pro Asp Leu Pro Glu Asp Lys Leu Ala Gln Ser Val Gln Lys Leu Gln
    710                 715                 720

GAC TTG ACA CTC CGA GAC CTG GAG AAG CAG GAA CGG GAA AAA GCT GCC      2322
Asp Leu Thr Leu Arg Asp Leu Glu Lys Gln Glu Arg Glu Lys Ala Ala
725                 730                 735                 740

AAC AGC TTG GAA GCG TTC ATA TTT GAG ACC CAG GAC AAG CTG TAC CAG      2370
Asn Ser Leu Glu Ala Phe Ile Phe Glu Thr Gln Asp Lys Leu Tyr Gln
                745                 750                 755

CCC GAG TAC CAG GAA GTG TCC ACA GAG GAG CAG CGT GAG GAG ATC TCT      2418
Pro Glu Tyr Gln Glu Val Ser Thr Glu Glu Gln Arg Glu Glu Ile Ser
            760                 765                 770

GGG AAG CTC AGC GCC GCA TCC ACC TGG CTG GAG GAT GAG GGT GTT GGA      2466
Gly Lys Leu Ser Ala Ala Ser Thr Trp Leu Glu Asp Glu Gly Val Gly
        775                 780                 785

GCC ACC ACA GTG ATG TTG AAG GAG AAG CTG GCT GAG CTG AGG AAG CTG      2514
Ala Thr Thr Val Met Leu Lys Glu Lys Leu Ala Glu Leu Arg Lys Leu
    790                 795                 800

TGC CAA GGG CTG TTT TTT CGG GTA GAG GAG CGC AAG AAG TGG CCC GAA      2562
Cys Gln Gly Leu Phe Phe Arg Val Glu Glu Arg Lys Lys Trp Pro Glu
805                 810                 815                 820

CGG CTG TCT GCC CTC GAT AAT CTC CTC AAC CAT TCC AGC ATG TTC CTC      2610
Arg Leu Ser Ala Leu Asp Asn Leu Leu Asn His Ser Ser Met Phe Leu
                825                 830                 835

AAG GGG GCC CGG CTC ATC CCA GAG ATG GAC CAG ATC TTC ACT GAG GTG      2658
Lys Gly Ala Arg Leu Ile Pro Glu Met Asp Gln Ile Phe Thr Glu Val
            840                 845                 850

GAG ATG ACA ACG TTA GAG AAA GTC ATC AAT GAG ACC TGG GCC TGG AAG      2706
Glu Met Thr Thr Leu Glu Lys Val Ile Asn Glu Thr Trp Ala Trp Lys
        855                 860                 865

AAT GCA ACT CTG GCC GAG CAG GCT AAG CTG CCC GCC ACA GAG AAG CCT      2754
Asn Ala Thr Leu Ala Glu Gln Ala Lys Leu Pro Ala Thr Glu Lys Pro
    870                 875                 880

GTG TTG CTC TCA AAA GAC ATT GAA GCT AAG ATG ATG GCC CTG GAC CGA      2802
Val Leu Leu Ser Lys Asp Ile Glu Ala Lys Met Met Ala Leu Asp Arg
885                 890                 895                 900
```

```
GAG GTG CAG TAT CTG CTC AAT AAG GCC AAG TTT ACC AAG CCC CGG CCC      2850
Glu Val Gln Tyr Leu Leu Asn Lys Ala Lys Phe Thr Lys Pro Arg Pro
                905                 910                 915

CGG CCT AAG GAC AAG AAT GGG ACC CGG GCA GAG CCA CCC CTC AAT GCC      2898
Arg Pro Lys Asp Lys Asn Gly Thr Arg Ala Glu Pro Pro Leu Asn Ala
            920                 925                 930

AGT GCC AGT GAC CAG GGG GAG AAG GTC ATC CCT CCA GCA GGC CAG ACT      2946
Ser Ala Ser Asp Gln Gly Glu Lys Val Ile Pro Pro Ala Gly Gln Thr
        935                 940                 945

GAA GAT GCA GAG CCC ATT TCA GAA CCT GAG AAA GTA GAG ACT GGA TCC      2994
Glu Asp Ala Glu Pro Ile Ser Glu Pro Glu Lys Val Glu Thr Gly Ser
    950                 955                 960

GAG CCA GGA GAC ACT GAG CCT TTG GAG TTA GGA GGT CCT GGA GCA GAA      3042
Glu Pro Gly Asp Thr Glu Pro Leu Glu Leu Gly Gly Pro Gly Ala Glu
965                 970                 975                 980

CCT GAA CAG AAA GAA CAA TCG ACA GGA CAG AAG CGG CCT TTG AAG AAC      3090
Pro Glu Gln Lys Glu Gln Ser Thr Gly Gln Lys Arg Pro Leu Lys Asn
                985                 990                 995

GAC GAA CTA TAACCCCCAC CTCTGTTTTC CCCATTCATC TCCACCCCCT               3139
Asp Glu Leu

TCCCCCACCA CTTCTATTTA TTTAACATCG AGGGTTGGGG GAGGGGTTGG TCCTGCCCTC    3199

GGCTGGAGTT CCTTTCTCAC CCCTGTGATT TGGAGGTGTG GAGAAGGGGA AGGGAGGGAC    3259

AGCTCACTGG TTCCTTCTGC AGTACCTCTG TGGTTAAAAA TGGAAACTGT TCTCCTCCCC    3319

AGCCCCACTC CCTGTTCCCT ACCCATATAG GCCCTAAATT TGGGAAAAAT CACTATTAAT    3379

TTCTGAATCC TTTGCCTGTG GGTAGGAAGA GAATGGCTGC CAGTGGCTGA TGGGTCCCGG    3439

TGATGGGAAG GGTATCAGGT TGCTGGGGAG TTTCCACTCT TCTCTGGTGA TTGTTCCTTC    3499

CCTCCCTTCC TCTCCCACCA TGCGATGAGC ATCCTTTCAG GCCAGTGTCT GCAGAGCCTC    3559

AGTTACCAGG TTTGGTTTCT GAGTGCCTAT CTGTGCTCTT TCCTCCCTCT GCGGGCTTCT    3619

CTTGCTCTGA GCCTCCCTTC CCCATTCCCA TGCAGCTCCT TTCCCCCTGG GTTTCCTTGG    3679

CTTCCTGCAG CAAATTGGGC AGTTCTCTGC CCCTTGCCTA AAAGCCTGTA CCTCTGGATT    3739

GGCGGAAGTA AATCTGGAAG GATTCTCACT CGTATTTCCC ACCCCTAGTG GCCAGAGGAG    3799

GGAGGGGCAC AGTGAAGAAG GGAGCCCACC ACCTCTCCGA AGAGGAAAGC CACGTAGAGT    3859

GGTTGGCATG GGGTGCCAGC ATCGTGCAAG CTCTGTCATA ATCTGCATCT TCCCAGCAGC    3919

CTGGTACCCC AGGTTCCTGT AACTCCCTGC CTCCTCCTCT CTTCTGCTGT TCTGCTCCTC    3979

CCAGACAGAG CCTTTCCCTC ACCCCCTGAC CCCCTGGGCT GACCAAAATG TGCTTTCTAC    4039

TGTGAGTCCC TATCCCAAGA TCCTGGGGAA AGGAGAGACC ATGGTGTGAA TGTAGAGATG    4099
```

CCACCTCCCT CTCTCTGAGG CAGGCCTGTG GATGAAGGAG GAGGGTCAGG GCTGGCCTTC 4159

CTCTGTGCAT CACTCTGCTA GGTTGGGGGC CCCCGACCCA CCATACCTAC GCCTAGGGAG 4219

CCCGTCCTCC AGTATTCCGT CTGTAGCAGG AGCTAGGGCT GCTGCCTCAG CTCCAAGACA 4279

AGAATGAACC TGGCTGTTGC AGTCATTTTG TCTTTTCCTT TTTTTTTTTT TGCCACATTG 4339

GCAGAGATGG GACCTAAGGG TCCCACCCCT CACCCCACCC CCACCTCTTC TGTATGTTTG 4399

AATTCTTTCA GTAGCTGTTG ATGCTGGTTG GACAGGTTTG AGTCAAATTG TACTTTGCTC 4459

CATTGTTAAT TGAGAAACTG TTTCAATAAA ATATTCTTTT CTAC 4503

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Ala Ala Thr Val Arg Arg Gln Arg Pro Arg Arg Leu Leu Cys Trp
1 5 10 15

Ala Leu Val Ala Val Leu Leu Ala Asp Leu Leu Ala Leu Ser Asp Thr
20 25 30

Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
35 40 45

Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu Ser
50 55 60

Arg Arg Lys Thr Pro Val Thr Val Thr Leu Lys Glu Asn Glu Arg Phe
65 70 75 80

Leu Gly Asp Ser Ala Ala Gly Met Ala Ile Lys Asn Pro Lys Ala Thr
85 90 95

Leu Arg Tyr Phe Gln His Leu Leu Gly Lys Gln Ala Asp Asn Pro His
100 105 110

Val Ala Leu Tyr Arg Ser Arg Phe Pro Glu His Glu Leu Asn Val Asp
115 120 125

Pro Gln Arg Gln Thr Val Arg Phe Gln Ile Ser Pro Gln Leu Gln Phe
130 135 140

Ser Pro Glu Glu Val Leu Gly Met Val Leu Asn Tyr Ser Arg Ser Leu
145 150 155 160

Ala Glu Asp Phe Ala Glu Gln Pro Ile Lys Asp Ala Val Ile Thr Val
165 170 175

Pro Ala Phe Phe Asn Gln Ala Glu Arg Arg Ala Val Leu Gln Ala Ala
180 185 190

```
Arg Met Ala Gly Leu Lys Val Leu Gln Leu Ile Asn Asp Asn Thr Ala
        195                 200                 205

Thr Ala Leu Ser Tyr Gly Val Phe Arg Arg Lys Asp Ile Asn Ser Thr
    210                 215                 220

Ala Gln Asn Ile Met Phe Tyr Asp Met Gly Ser Gly Ser Thr Val Cys
225                 230                 235                 240

Thr Ile Val Thr Tyr Gln Thr Val Lys Thr Lys Glu Ala Gly Thr Gln
                245                 250                 255

Pro Gln Leu Gln Ile Arg Gly Val Gly Phe Asp Arg Thr Leu Gly Gly
            260                 265                 270

Leu Glu Met Glu Leu Arg Leu Arg Glu His Leu Ala Lys Leu Phe Asn
        275                 280                 285

Glu Gln Arg Lys Gly Gln Lys Ala Lys Asp Val Arg Glu Asn Pro Arg
    290                 295                 300

Ala Met Ala Lys Leu Leu Arg Glu Ala Asn Arg Leu Lys Thr Val Leu
305                 310                 315                 320

Ser Ala Asn Ala Asp His Met Ala Gln Ile Glu Gly Leu Met Asp Asp
                325                 330                 335

Val Asp Phe Lys Ala Lys Val Thr Arg Val Glu Phe Glu Glu Leu Cys
            340                 345                 350

Ala Asp Leu Phe Asp Arg Val Pro Gly Pro Val Gln Gln Ala Leu Gln
        355                 360                 365

Ser Ala Glu Met Ser Leu Asp Gln Ile Glu Gln Val Ile Leu Val Gly
    370                 375                 380

Gly Pro Thr Arg Val Pro Lys Val Gln Glu Val Leu Leu Lys Pro Val
385                 390                 395                 400

Gly Lys Glu Glu Leu Gly Lys Asn Ile Asn Ala Asp Glu Ala Ala Ala
                405                 410                 415

Met Gly Ala Val Tyr Gln Ala Ala Ala Leu Ser Lys Ala Phe Lys Val
            420                 425                 430

Lys Pro Phe Val Val Arg Asp Ala Val Ile Tyr Pro Ile Leu Val Glu
        435                 440                 445

Phe Thr Arg Glu Val Glu Glu Glu Pro Gly Leu Arg Ser Leu Lys His
    450                 455                 460

Asn Lys Arg Val Leu Phe Ser Arg Met Gly Pro Tyr Pro Gln Arg Lys
465                 470                 475                 480

Val Ile Thr Phe Asn Arg Tyr Ser His Asp Phe Asn Phe His Ile Asn
                485                 490                 495

Tyr Gly Asp Leu Gly Phe Leu Gly Pro Glu Asp Leu Arg Val Phe Gly
            500                 505                 510
```

Ser Gln Asn Leu Thr Thr Val Lys Leu Lys Gly Val Gly Glu Ser Phe
515 520 525

Lys Lys Tyr Pro Asp Tyr Glu Ser Lys Gly Ile Lys Ala His Phe Asn
530 535 540

Leu Asp Glu Ser Gly Val Leu Ser Leu Asp Arg Val Glu Ser Val Phe
545 550 555 560

Glu Thr Leu Val Glu Asp Ser Pro Glu Glu Glu Ser Thr Leu Thr Lys
565 570 575

Leu Gly Asn Thr Ile Ser Ser Leu Phe Gly Gly Gly Thr Ser Ser Asp
580 585 590

Ala Lys Glu Asn Gly Thr Asp Ala Val Gln Glu Glu Glu Glu Ser Pro
595 600 605

Ala Glu Gly Ser Lys Asp Glu Pro Ala Glu Gln Gly Glu Leu Lys Glu
610 615 620

Glu Ala Glu Ala Pro Met Glu Asp Thr Ser Gln Pro Pro Pro Ser Glu
625 630 635 640

Pro Lys Gly Asp Ala Ala Arg Glu Gly Glu Thr Pro Asp Glu Lys Glu
645 650 655

Ser Gly Asp Lys Ser Glu Ala Gln Lys Pro Asn Glu Lys Gly Gln Ala
660 665 670

Gly Pro Glu Gly Val Pro Pro Ala Pro Glu Glu Glu Lys Lys Gln Lys
675 680 685

Pro Ala Arg Lys Gln Lys Met Val Glu Glu Ile Gly Val Glu Leu Ala
690 695 700

Val Leu Asp Leu Pro Asp Leu Pro Glu Asp Glu Leu Ala His Ser Val
705 710 715 720

Gln Lys Leu Glu Asp Leu Thr Leu Arg Asp Leu Glu Lys Gln Glu Arg
725 730 735

Glu Lys Ala Ala Asn Ser Leu Glu Ala Phe Ile Phe Glu Thr Gln Asp
740 745 750

Lys Leu Tyr Gln Pro Glu Tyr Gln Glu Val Ser Thr Glu Glu Gln Arg
755 760 765

Glu Glu Ile Ser Gly Lys Leu Ser Ala Thr Ser Thr Trp Leu Glu Asp
770 775 780

Glu Gly Phe Gly Ala Thr Thr Val Met Leu Lys Asp Lys Leu Ala Glu
785 790 795 800

Leu Arg Lys Leu Cys Gln Gly Leu Phe Phe Arg Val Glu Glu Arg Arg
805 810 815

Lys Trp Pro Glu Arg Leu Ser Ala Leu Asp Asn Leu Leu Asn His Ser
820 825 830

```
Ser Ile Phe Leu Lys Gly Ala Arg Leu Ile Pro Glu Met Asp Gln Ile
        835                 840                 845

Phe Thr Asp Val Glu Met Thr Thr Leu Glu Lys Val Ile Asn Asp Thr
    850                 855                 860

Trp Thr Trp Lys Asn Ala Thr Leu Ala Glu Gln Ala Lys Leu Pro Ala
865                 870                 875                 880

Thr Glu Lys Pro Val Leu Leu Ser Lys Asp Ile Glu Ala Lys Met Met
                885                 890                 895

Ala Leu Asp Arg Glu Val Gln Tyr Leu Leu Asn Lys Ala Lys Phe Thr
            900                 905                 910

Lys Pro Arg Pro Arg Pro Lys Asp Lys Asn Gly Thr Arg Thr Glu Pro
        915                 920                 925

Pro Leu Asn Ala Ser Ala Gly Asp Gln Glu Glu Lys Val Ile Pro Pro
    930                 935                 940

Thr Gly Gln Thr Glu Glu Ala Lys Ala Ile Leu Glu Pro Asp Lys Glu
945                 950                 955                 960

Gly Leu Gly Thr Glu Ala Ala Asp Ser Glu Pro Leu Glu Leu Gly Gly
                965                 970                 975

Pro Gly Ala Glu Ser Glu Gln Ala Glu Gln Thr Ala Gly Gln Lys Arg
            980                 985                 990

Pro Leu Lys Asn Asp Glu Leu
        995
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3252 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION:203..3199

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
TGAGGATGGA GCAGCGGTCG GGCCGCGGCT CCTAGGGGAG GCAGCGTGCT AGCTTCGGGG      60

GCGGGCCAGT AGCGGGAGCG AGGGCCGTAC GGACACCGGT CCCTTCGGCC TTGAAGTTCA     120

GGCGCTGAGC TGCCCCCTCG CGCTCGGGGT GGGCCGGAAT CCATTTCTGG GAGTGGGATC     180
```

```
TTCCACCTTC ATCAGGGTCA CA ATG GCA GCT ACA GTA AGG AGG CAG AGG CCA      232
                         Met Ala Ala Thr Val Arg Arg Gln Arg Pro
                                                5                  10

AGG AGG CTA CTC TGT TGG GCC TTG GTG GCT GTC CTC TTG GCA GAC CTG      280
Arg Arg Leu Leu Cys Trp Ala Leu Val Ala Val Leu Leu Ala Asp Leu
                    15                  20                  25

TTG GCA CTG AGT GAC ACA CTG GCT GTG ATG TCT GTG GAC CTG GGC AGT      328
Leu Ala Leu Ser Asp Thr Leu Ala Val Met Ser Val Asp Leu Gly Ser
                30                  35                  40

GAA TCC ATG AAG GTG GCC ATT GTC AAG CCT GGA GTG CCC ATG GAG ATT      376
Glu Ser Met Lys Val Ala Ile Val Lys Pro Gly Val Pro Met Glu Ile
            45                  50                  55

GTA TTG AAC AAG GAA TCT CGG AGG AAA ACT CCG GTG ACT GTG ACC TTG      424
Val Leu Asn Lys Glu Ser Arg Arg Lys Thr Pro Val Thr Val Thr Leu
        60                  65                  70

AAG GAA AAC GAA AGG TTT CTA GGT GAC AGT GCA GCT GGC ATG GCC ATC      472
Lys Glu Asn Glu Arg Phe Leu Gly Asp Ser Ala Ala Gly Met Ala Ile
 75                  80                  85                  90

AAG AAC CCA AAG GCT ACG CTC CGT TAT TTC CAG CAC CTC CTT GGA AAG      520
Lys Asn Pro Lys Ala Thr Leu Arg Tyr Phe Gln His Leu Leu Gly Lys
                95                  100                 105

CAG GCA GAT AAC CCT CAT GTG GCT CTT TAC CGG TCC CGT TTC CCA GAA      568
Gln Ala Asp Asn Pro His Val Ala Leu Tyr Arg Ser Arg Phe Pro Glu
            110                 115                 120

CAT GAG CTC AAT GTT GAC CCA CAG AGG CAG ACT GTG CGC TTC CAG ATC      616
His Glu Leu Asn Val Asp Pro Gln Arg Gln Thr Val Arg Phe Gln Ile
        125                 130                 135

AGT CCG CAG CTG CAG TTC TCT CCC GAG GAG GTG CTG GGC ATG GTT CTC      664
Ser Pro Gln Leu Gln Phe Ser Pro Glu Glu Val Leu Gly Met Val Leu
    140                 145                 150

AAC TAC TCC CGT TCC CTG GCT GAA GAT TTT GCA GAA CAA CCT ATT AAG      712
Asn Tyr Ser Arg Ser Leu Ala Glu Asp Phe Ala Glu Gln Pro Ile Lys
155                 160                 165                 170

GAT GCA GTG ATC ACC GTG CCA GCC TTT TTC AAC CAG GCC GAG CGC CGA      760
Asp Ala Val Ile Thr Val Pro Ala Phe Phe Asn Gln Ala Glu Arg Arg
                175                 180                 185

GCT GTG CTG CAG GCT GCT CGT ATG GCT GGC CTC AAG GTG CTG CAG CTC      808
Ala Val Leu Gln Ala Ala Arg Met Ala Gly Leu Lys Val Leu Gln Leu
            190                 195                 200

ATC AAT GAC AAC ACT GCC ACA GCC CTC AGC TAT GGT GTC TTC CGC CGG      856
Ile Asn Asp Asn Thr Ala Thr Ala Leu Ser Tyr Gly Val Phe Arg Arg
        205                 210                 215

AAA GAT ATC AAT TCC ACT GCA CAG AAT ATC ATG TTC TAT GAC ATG GGC      904
Lys Asp Ile Asn Ser Thr Ala Gln Asn Ile Met Phe Tyr Asp Met Gly
    220                 225                 230
```

```
TCG GGC AGC ACT GTG TGT ACC ATC GTG ACC TAC CAA ACG GTG AAG ACT      952
Ser Gly Ser Thr Val Cys Thr Ile Val Thr Tyr Gln Thr Val Lys Thr
235                 240                 245                 250

AAG GAG GCT GGG ACG CAG CCA CAG CTA CAG ATC CGG GGC GTG GGA TTT     1000
Lys Glu Ala Gly Thr Gln Pro Gln Leu Gln Ile Arg Gly Val Gly Phe
                255                 260                 265

GAC CGC ACC CTG GGT GGC CTG GAG ATG GAG CTT CGG CTG CGA GAG CAC     1048
Asp Arg Thr Leu Gly Gly Leu Glu Met Glu Leu Arg Leu Arg Glu His
            270                 275                 280

CTG GCT AAG CTC TTC AAT GAG CAG CGC AAG GGC CAG AAA GCC AAG GAT     1096
Leu Ala Lys Leu Phe Asn Glu Gln Arg Lys Gly Gln Lys Ala Lys Asp
        285                 290                 295

GTT CGG GAA AAC CCC CGA GCC ATG GCC AAA CTG CTT CGG GAA GCC AAT     1144
Val Arg Glu Asn Pro Arg Ala Met Ala Lys Leu Leu Arg Glu Ala Asn
    300                 305                 310

CGG CTT AAA ACC GTC CTG AGT GCC AAT GCT GAT CAC ATG GCA CAG ATT     1192
Arg Leu Lys Thr Val Leu Ser Ala Asn Ala Asp His Met Ala Gln Ile
315                 320                 325                 330

GAA GGC TTG ATG GAC GAT GTG GAC TTC AAG GCA AAA GTA ACT CGA GTG     1240
Glu Gly Leu Met Asp Asp Val Asp Phe Lys Ala Lys Val Thr Arg Val
                335                 340                 345

GAG TTT GAG GAG CTG TGT GCA GAT TTG TTT GAT CGA GTG CCT GGG CCT     1288
Glu Phe Glu Glu Leu Cys Ala Asp Leu Phe Asp Arg Val Pro Gly Pro
            350                 355                 360

GTA CAG CAG GCC CTG CAG AGT GCT GAG ATG AGC CTG GAT CAA ATT GAG     1336
Val Gln Gln Ala Leu Gln Ser Ala Glu Met Ser Leu Asp Gln Ile Glu
        365                 370                 375

CAG GTG ATC CTG GTG GGT GGG CCC ACT CGT GTT CCC AAA GTT CAA GAG     1384
Gln Val Ile Leu Val Gly Gly Pro Thr Arg Val Pro Lys Val Gln Glu
    380                 385                 390

GTG CTG CTG AAG CCT GTG GGC AAG GAG GAA CTA GGA AAG AAC ATC AAT     1432
Val Leu Leu Lys Pro Val Gly Lys Glu Glu Leu Gly Lys Asn Ile Asn
395                 400                 405                 410

GCC GAT GAA GCA GCT GCC ATG GGG GCC GTG TAC CAG GCA GCG GCA CTG     1480
Ala Asp Glu Ala Ala Ala Met Gly Ala Val Tyr Gln Ala Ala Ala Leu
                415                 420                 425

AGC AAA GCC TTC AAA GTG AAG CCA TTT GTT GTG CGT GAT GCT GTT ATT     1528
Ser Lys Ala Phe Lys Val Lys Pro Phe Val Val Arg Asp Ala Val Ile
            430                 435                 440

TAC CCC ATC CTG GTG GAG TTC ACA AGG GAG GTG GAG GAG GAG CCT GGG     1576
Tyr Pro Ile Leu Val Glu Phe Thr Arg Glu Val Glu Glu Glu Pro Gly
        445                 450                 455

CTT CGA AGC CTG AAG CAC AAT AAA CGT GTG CTC TTC TCC CGA ATG GGG     1624
Leu Arg Ser Leu Lys His Asn Lys Arg Val Leu Phe Ser Arg Met Gly
    460                 465                 470
```

```
CCC TAC CCT CAG CGC AAA GTC ATC ACC TTT AAC CGA TAC AGC CAT GAT      1672
Pro Tyr Pro Gln Arg Lys Val Ile Thr Phe Asn Arg Tyr Ser His Asp
475                 480                 485                 490

TTC AAC TTT CAC ATC AAC TAC GGT GAC CTG GGC TTC CTG GGG CCT GAG      1720
Phe Asn Phe His Ile Asn Tyr Gly Asp Leu Gly Phe Leu Gly Pro Glu
                495                 500                 505

GAT CTT CGG GTA TTT GGC TCC CAG AAT CTG ACC ACA GTG AAA CTA AAA      1768
Asp Leu Arg Val Phe Gly Ser Gln Asn Leu Thr Thr Val Lys Leu Lys
            510                 515                 520

GGT GTG GGA GAG AGC TTC AAG AAA TAT CCT GAC TAT GAG TCC AAA GGC      1816
Gly Val Gly Glu Ser Phe Lys Lys Tyr Pro Asp Tyr Glu Ser Lys Gly
        525                 530                 535

ATC AAG GCC CAC TTT AAC CTA GAC GAG AGT GGA GTG CTC AGT TTA GAC      1864
Ile Lys Ala His Phe Asn Leu Asp Glu Ser Gly Val Leu Ser Leu Asp
    540                 545                 550

AGG GTG GAG TCC GTA TTC GAG ACC CTG GTG GAG GAC AGC CCA GAG GAA      1912
Arg Val Glu Ser Val Phe Glu Thr Leu Val Glu Asp Ser Pro Glu Glu
555                 560                 565                 570

GAG TCT ACT CTT ACC AAA CTT GGC AAC ACC ATT TCC AGC CTG TTT GGC      1960
Glu Ser Thr Leu Thr Lys Leu Gly Asn Thr Ile Ser Ser Leu Phe Gly
                575                 580                 585

GGT GGT ACC TCA TCA GAT GCC AAA GAG AAT GGT ACT GAT GCT GTA CAG      2008
Gly Gly Thr Ser Ser Asp Ala Lys Glu Asn Gly Thr Asp Ala Val Gln
            590                 595                 600

GAG GAG GAG GAG AGC CCT GCT GAG GGG AGC AAG GAT GAG CCT GCA GAA      2056
Glu Glu Glu Glu Ser Pro Ala Glu Gly Ser Lys Asp Glu Pro Ala Glu
        605                 610                 615

CAG GGG GAA CTC AAG GAG GAA GCT GAA GCC CCA ATG GAG GAT ACC TCC      2104
Gln Gly Glu Leu Lys Glu Glu Ala Glu Ala Pro Met Glu Asp Thr Ser
    620                 625                 630

CAG CCT CCA CCC TCT GAG CCT AAG GGG GAT GCA GCC CGT GAG GGA GAA      2152
Gln Pro Pro Pro Ser Glu Pro Lys Gly Asp Ala Ala Arg Glu Gly Glu
635                 640                 645                 650

ACA CCT GAT GAA AAA GAA AGT GGG GAC AAG TCT GAG GCC CAG AAG CCC      2200
Thr Pro Asp Glu Lys Glu Ser Gly Asp Lys Ser Glu Ala Gln Lys Pro
                655                 660                 665

AAT GAG AAG GGG CAG GCA GGG CCT GAG GGT GTC CCT CCA GCT CCC GAG      2248
Asn Glu Lys Gly Gln Ala Gly Pro Glu Gly Val Pro Pro Ala Pro Glu
            670                 675                 680

GAA GAA AAA AAG CAG AAA CCT GCC CGG AAG CAG AAA ATG GTG GAG GAG      2296
Glu Glu Lys Lys Gln Lys Pro Ala Arg Lys Gln Lys Met Val Glu Glu
        685                 690                 695

ATA GGT GTG GAA CTG GCT GTC TTG GAC CTG CCA GAC TTG CCA GAG GAT      2344
Ile Gly Val Glu Leu Ala Val Leu Asp Leu Pro Asp Leu Pro Glu Asp
    700                 705                 710
```

```
GAG CTG GCC CAT TCC GTG CAG AAA CTT GAG GAC TTG ACC CTG CGA GAC      2392
Glu Leu Ala His Ser Val Gln Lys Leu Glu Asp Leu Thr Leu Arg Asp
715                 720                 725                 730

CTT GAA AAG CAG GAG AGG GAG AAA GCT GCC AAC AGC TTA GAA GCT TTT      2440
Leu Glu Lys Gln Glu Arg Glu Lys Ala Ala Asn Ser Leu Glu Ala Phe
                735                 740                 745

ATC TTT GAG ACC CAG GAC AAA CTG TAC CAA CCT GAG TAC CAG GAA GTG      2488
Ile Phe Glu Thr Gln Asp Lys Leu Tyr Gln Pro Glu Tyr Gln Glu Val
            750                 755                 760

TCC ACT GAG GAA CAA CGG GAG GAG ATC TCT GGA AAA CTC AGT GCC ACT      2536
Ser Thr Glu Glu Gln Arg Glu Glu Ile Ser Gly Lys Leu Ser Ala Thr
        765                 770                 775

TCT ACC TGG CTG GAG GAT GAG GGA TTT GGA GCC ACC ACT GTG ATG TTG      2584
Ser Thr Trp Leu Glu Asp Glu Gly Phe Gly Ala Thr Thr Val Met Leu
    780                 785                 790

AAG GAC AAG CTG GCT GAG CTG AGA AAG CTG TGC CAA GGG CTG TTT TTT      2632
Lys Asp Lys Leu Ala Glu Leu Arg Lys Leu Cys Gln Gly Leu Phe Phe
795                 800                 805                 810

CGG GTG GAA GAG CGC AGG AAA TGG CCA GAG CGG CTT TCA GCT CTG GAT      2680
Arg Val Glu Glu Arg Arg Lys Trp Pro Glu Arg Leu Ser Ala Leu Asp
                815                 820                 825

AAT CTC CTC AAT CAC TCC AGC ATT TTC CTC AAG GGT GCC CGA CTC ATC      2728
Asn Leu Leu Asn His Ser Ser Ile Phe Leu Lys Gly Ala Arg Leu Ile
            830                 835                 840

CCA GAG ATG GAC CAG ATC TTC ACT GAC GTG GAG ATG ACA ACG TTG GAG      2776
Pro Glu Met Asp Gln Ile Phe Thr Asp Val Glu Met Thr Thr Leu Glu
        845                 850                 855

AAA GTC ATC AAT GAC ACC TGG ACC TGG AAG AAT GCA ACC CTG GCC GAG      2824
Lys Val Ile Asn Asp Thr Trp Thr Trp Lys Asn Ala Thr Leu Ala Glu
    860                 865                 870

CAG GCC AAG CTT CCT GCC ACA GAG AAA CCC GTG CTG CTT TCA AAA GAC      2872
Gln Ala Lys Leu Pro Ala Thr Glu Lys Pro Val Leu Leu Ser Lys Asp
875                 880                 885                 890

ATC GAG GCC AAA ATG ATG GCC CTG GAC CGG GAG GTG CAG TAT CTA CTC      2920
Ile Glu Ala Lys Met Met Ala Leu Asp Arg Glu Val Gln Tyr Leu Leu
                895                 900                 905

AAT AAG GCC AAG TTT ACT AAA CCC CGG CCA CGG CCC AAG GAC AAG AAT      2968
Asn Lys Ala Lys Phe Thr Lys Pro Arg Pro Arg Pro Lys Asp Lys Asn
            910                 915                 920

GGC ACC CGG ACA GAG CCT CCC CTC AAT GCC AGT GCT GGT GAC CAA GAG      3016
Gly Thr Arg Thr Glu Pro Pro Leu Asn Ala Ser Ala Gly Asp Gln Glu
        925                 930                 935

GAA AAG GTC ATT CCA CCT ACA GGC CAG ACT GAA GAG GCG AAG GCC ATC      3064
Glu Lys Val Ile Pro Pro Thr Gly Gln Thr Glu Glu Ala Lys Ala Ile
    940                 945                 950
```

```
TTA GAA CCT GAC AAA GAA GGG CTT GGT ACA GAG GCA GCA GAC TCT GAG      3112
Leu Glu Pro Asp Lys Glu Gly Leu Gly Thr Glu Ala Ala Asp Ser Glu
955                 960                 965                 970

CCT CTG GAA TTA GGA GGT CCT GGT GCA GAA TCT GAA CAG GCA GAG CAG      3160
Pro Leu Glu Leu Gly Gly Pro Gly Ala Glu Ser Glu Gln Ala Glu Gln
                975                 980                 985

ACA GCA GGG CAG AAG CGG CCT TTG AAG AAT GAT GAG CTG TGACCCCGCG      3209
Thr Ala Gly Gln Lys Arg Pro Leu Lys Asn Asp Glu Leu
            990                 995

CCTCCGCTCC ACTTGCCTCC AGCCCCTTCT CCTACCACCT CTA                     3252
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Leu Ala Val Met Ser Val Asp Leu Gly Ser Glu Ser Met Lys Val Ala
1               5                   10                  15

Ile Val Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic nucleic acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
AATACGACTC ACTATAGGGA                                              20
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Lys Pro Gly Val Pro Met Glu
1 5

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic nucleic acid"

(ix) FEATURE:
(A) NAME/KEY: -
(B) LOCATION:6
(D) OTHER INFORMATION:/note= "N at position 6 is an inosine residue."

(ix) FEATURE:
(A) NAME/KEY: -
(B) LOCATION:9
(D) OTHER INFORMATION:/note= "N at position 9 is an inosine residue."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AARCCNGGNG TNCCNATGGA 20

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Lys Pro Gly Val Pro Met Glu Ile Val Leu Asn Lys Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic nucleic acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GCACCCTTGA GGAAAATGCT 20

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic nucleic acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCAGAAGCC CAATGAGAAG 20

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2861 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GAAAGAAGTA GACATGGGAG ACTTCATTTT GTTCTGTACT AAGAAAAATT CTTCTGCCTT 60

GGGATGCTGT TGATCTATGA CCTTACCCCC AACCCTGTGC TCTCTGAAAC ATGTGCTGTG 120

TCCACTCAGG GTTAAATGGA TTAAGGGCGG TGCAAGATGT GCTTTGTTAA ACAGATGCTT 180

GAAGGCAGCA TGCTCGTTAG GAGTCATCAC CACTCCCTAA TCTCAAGTAC CCAGGGACAC 240

AAACACTGCG GAAGGCCACA GGGTCCTCTG CCTAGGAAAG CCAGAGACCT TTGTTCACTT 300

GTTTATCTGC TGACCTTCCC TCCACTATTG TCCTATGACC CTGCCAAATC CCCCTCTGCC 360

AGAAACACCC AAGAATGATC AATAAAAAAA AAAAAAAAAA AAAAAGGAAG AATAGACTCT 420

CTCTGGGACT GCCAATAATT TTTCCTTCTA AGCATAGACA CCGGACCACT CTCCACCTAA 480

GCATCACGAA AAATGTAGAG AAAGGAAGAG CTAAGAGCTC CTTAAACAAG TTCAGGCTTG 540

ACACAACCCT GGCCCTGACA GCCAGGGTCT TCAAGCGGGC CTTTCTGTGA AGGGTGGCCA 600

GGCATCAACT TAGTAGGAGA GAAAACAGAT GACTTATTTC CATCCACACT TAAGGAAAAT 660

GCAGTCTCCA AGGACTGCGT ACATTTCTTT TTCGAGAAGG AGTCTCGCTG TTGTCGCCCA 720

GGCTGGAGTG CAGTGGCGCA GTCTGGGCTC ACAGCAACCT CTGCCTCCCG GATTCAAGCA 780

ATTCTCCTGC CTCAGCCTCG TGAGTAGCTG GGATTACAGG CACCCGCCAC CACGCCTGGC 840

TAATTTTTGT AGTTTTGGTA GAGACGGGGT TTCACCATGT TGGCCAGGCT GGTCTCGAAC 900

TCCTGACCTC CAGTGATTCG CCCGCCTTGG CCTCCCAAAA TGCTGGGATT ACAGGCGTGA 960

GCCACCGCGC CCGGGCGACT GCGCACATTT CTATGGAGCT GTAAGTTAAA AGAGAAGGCA 1020

GTGAGGTGCT TCTGTCATTC TATGACAGAA ACAGCTAAAG AGTAGAGAAA TGTTCACAAG 1080

ATTTAATAGA ACAGAAATAG GAGAAGGTGC ACACAAGCTC AACCAACTAT AGCCTCACAA 1140

ATAAAAGTGT CTTTTGTGTG TAGTACTTAA GTTTGGAATA TTCTTTCTTA TACAAATGAG 1200

TGGGGCTTAA CCTAAGAAAT CCTGGCCAGA TTCTGCGACG AATGCATCGG TTATCTCTGA 1260

CCCATCAGCA AACATCTTTT TCTGTGGCTT CAGTTTCCTC AGTAAAACAG AGGGGGTTGC 1320

GACGGACTCA GTCCGAGGCA CAGCCATTCT CCAACGTCTA TCCAAAGCCT AGGGCACCTC 1380

AATACTAACC GGCAGGCCAG CGCCCCCTCC GCGGGGCTGC GGACAGGACG CCTGTTATTC 1440

CATTCCTCGG CCGGGCTCTA CAGGTGACCG GAAGAAGAGC CCCGAGTGCG GGACTGCAGT 1500

GCGCCCGACC TGCTCTAGGC GCAGGTCACT CCCGAACCCC GGCAGCAAAG CATCCAGCGC 1560

CGGAAAAGGT CCCGCGGTCG CCCCGGGGCC GGCGCTGGGG AGGAAGGAGT GGAGCGCGCT 1620

GGCCCCGTGA CGTGGTCCAA TCCCAGGCCG ACGCCGGCTG CTTCTGCCCA ACCGGTGGCT 1680

GGTCCCCTCC GCCGCCCCCA TTACAAGGCT GGCAAAGGGA GGGGGCGGGG CCTGGGACGT 1740

GGTCCAATGA GTACGCGCGC CGGGGCGGCG GGGGCGGGGC CGGGCGCGCA GCGCAGGGCC 1800

GGGCGGCCGA GGCTCCAATG AGCGCCCGCC GCGTCCGGGG CCGGCTGGTG CGCGAGACGC 1860

CGCCGAGAGG TTGGTGGCTA ATGTAACAGT TTGCAAACCG AGAGGAGTTG TGAAGGGCGC 1920

GGGTGGGGGG CGCTGCCGGC CTCGTGGGTA CGTTCGTGCC GCGTCTGTCC CAGAGCTGGG 1980

GCCGCAGGAG CGGAGGCAAG AGGTAGCGGG GGTGGATGGA GGTGCGGGCC GGCCACCCCT 2040

CCTAGGGGAG ACAGCGTGCG AGCTCCGGGG GCGGGTCGGG AGCGCAAGGG AGGGCCGCGC 2100

```
GGACGCCGGG CGCTCGGCCT CGCACCGGGG GGCACGCAGC TCGGCCCCCG GTCTGTCCCC    2160

ACTTGCTGGG GCGGGCCGGG ATCCGTTTCC GGGAGTGGGA GCCGCCGCCT TCGTCAGGTG    2220

GGGTTTAGGT GAACACCGGG TAACGGCTAC CCGCCGGGCG GGGAACCTTA CCGCCCCTGG    2280

CACTGCGTCT GTGGGCACAG CGGGGCCGGG GAGTGAGCTG GGAAAGGGGA GGGGGCGGGA    2340

CAACCCGCAG GGATGCCGAG GAGGAGATAG GCCTTTCCTT CATCCTAGCT ACCCCCAACG    2400

TCATTACCTT TCTCTTCCCG TCCAGGCCCA GCTGGCTTTC CCCGTCAGCG GGGGAGCTCC    2460

AGGTGTGGGG AGGTGGTTGA GCCCTGGGCG GGGATCCCTG GCCGCACCCC AGGTGTCTGA    2520

CAACAGGCAC AGTGCTGCGG TGCGCCACTC ACTGCCTGTG TGGTGGACAA AAGGCTCGGG    2580

TCTCCTTTCT CTTGTCCTGT TAGCTTCTCT GTTTAGGGAT GTGGCAAAGC CGAGGACCCA    2640

TGCTCTTTCA CTTGGGCCTT TGTGTGGGCG CTGCTGGGAT GATTAGAGAA TGGTTTGTAC    2700

CCATCAGGAG GGAGAAGGGG AGAAGTAGGC TGATCTGCCC TGGGTAAGAA TGAAGTAGAT    2760

ATGAATCTTA CAGCCTCTCC GTTCTGGGAT GTGATTCTGT CTCCTTCACT CCGGGTATCC    2820

AGTTTTAAGT GTTTTCTTTC TTCGCCTCCC CCAGGGGCAC T                        2861
```

Claims

1. A polynucleotide encoding an ORP150 polypeptide selected from the group consisting of:

   (a) polynucleotides encoding the polypeptide having the amino acid sequence as depicted in SEQ ID NO:1 or a fragment of the polypeptide;
   (b) polynucleotides comprising the coding region of the nucleotide sequence as shown in SEQ ID NO:2 or a fragment thereof;
   (c) polynucleotides encoding the polypeptide having the amino acid sequence as depicted in SEQ ID NO:3 or a fragment of the polypeptide;
   (d) polynucleotides comprising the coding region of the nucleotide sequence as depicted in SEQ ID NO:4 or a fragment thereof;
   (e) polynucleotides encoding an ORP150 polypeptide which differs from the polypeptide encoded by the polynucleotide of (a) or (c) due to deletion(s), addition(s), insertion(s) and/or substitutions(s) of one or more amino acid residues; and
   (f) polynucleotides the complementary strand of which hybridizes to a polynucleotide of any one of (a) to (e) and which encode an ORP150 polypeptide;

   and the complementary strand of such a polynucleotide.

2. The polynucleotide of claim 1 which is DNA.

3. The polynucleotide of claim 2 which is genomic DNA.

4. The polynucleotide of claim 1 which is RNA.

5. A vector comprising the polynucleotide of any one of claims 1 to 4.

6. The vector of claim 5, in which the polynucleotide is operatively linked to regulatory elements which allow for expression in prokaryotic or eukaryotic host cells.

7. A host cell transformed and genetically engineered with a polynucleotide of any one of claims 1 to 4 or with a vector of claim 5 or 6.

8. A process for the preparation of an ORP150 polypeptide comprising culturing the host cell of claim 7 and recovering the polypeptide from the cells and/or the culture medium.

9. A polypeptide encoded by the polynucleotide of any one of claims 1 to 4 or obtainable by the process of claim 8.

10. An antibody or fragment thereof which specifically recognizes the polypeptide of claim 9.

11. A nucleic acid molecule which specifically hybridizes to a polynucleotide of any one of claims 1 to 4.

12. A pharmaceutical composition comprising a polynucleotide of any one of claims 1 to 4, the polypeptide of claim 9, the antibody of claim 10 and/or the nucleic acid molecule of claim 11 and optionally a pharmaceutically acceptable carrier.

13. A diagnostic composition comprising a polynucleotide of any one of claims 1 to 4, the polypeptide of claim 9, the antibody of claim 10 and/or the nucleic acid molecule of claim 11.

14. Use of the polynucleotide of any one of claims 1 to 4, the polypeptide of claim 9, the antibody of claim 10 or the nucleic acid molecule of claim 11 for the preparation of a pharmaceutical composition for the treatment of ischemic diseases.

15. A nucleic acid molecule having promoter activity and being able to promote transcription in cells when exposed to hypoxia selected from the group consisting of:

(a) polynucleotides comprising the nucleotide sequence as depicted in SEQ ID NO:12 or a fragment thereof; and
(b) polynucleotides hybridizing with the polynucleotide of (a).

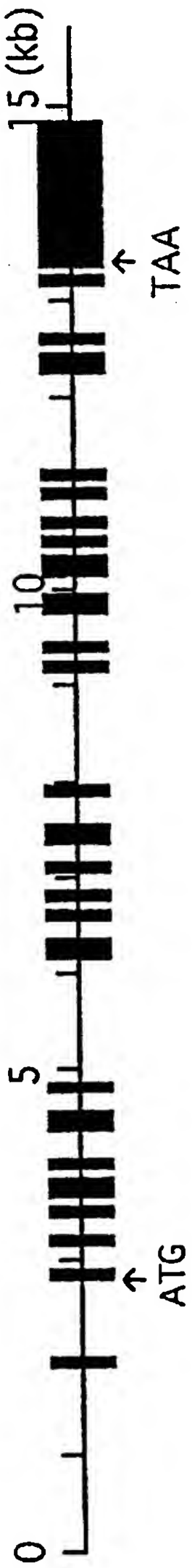

FIGURE 1

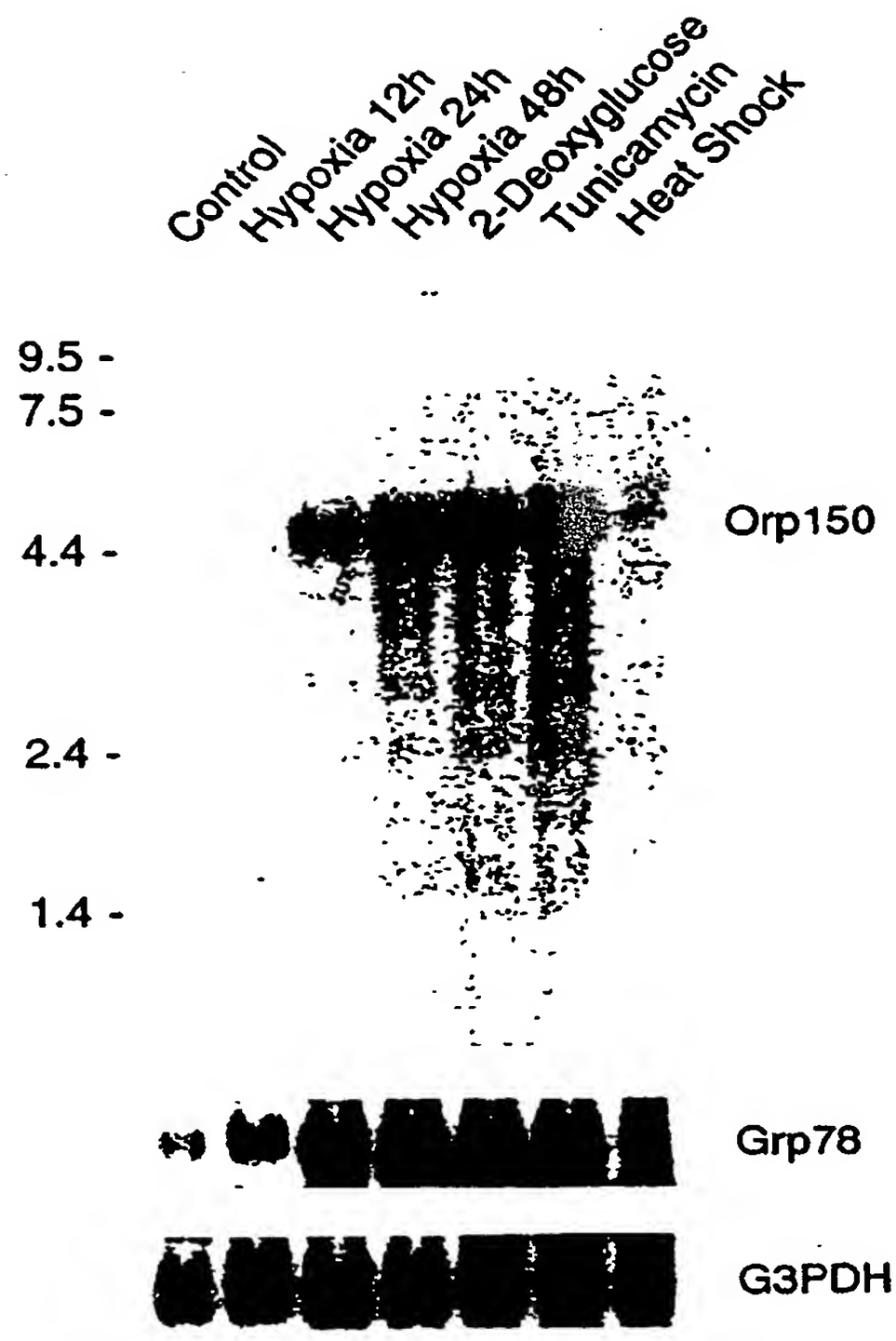

FIGURE 2

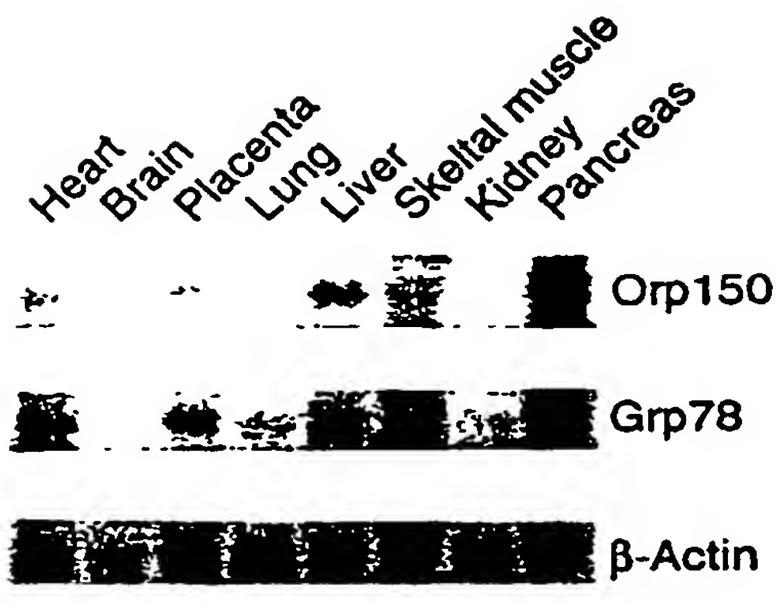
Heart
Brain
Placenta
Lung
Liver
Skeltal muscle
Kidney
Pancreas
Orp150
Grp78
β-Actin

FIGURE 3

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 780 472 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
20.05.1998 Bulletin 1998/21

(43) Date of publication A2:
25.06.1997 Bulletin 1997/26

(21) Application number: **96120662.0**

(22) Date of filing: **20.12.1996**

(51) Int. Cl.[6]: **C12N 15/12**, C07K 14/435, C12N 1/21, C12N 15/70, C07K 16/18, A61K 31/70, C12Q 1/68, A61K 39/00, G01N 33/577, C12N 15/79

(84) Designated Contracting States:
AT BE CH DE ES FR GB IT LI NL SE

(30) Priority: **20.12.1995 JP 349661/95**
23.07.1996 JP 213181/96

(71) Applicant:
Hsp Research Institute, Inc.
Chuo-ku, Osaka-shi, Osaka (JP)

(72) Inventors:
- **Ikeda, Jun**
  Tokyo (JP)
- **Kaneda, Sumiko**
  Kyoto-shi, Kyoto (JP)
- **Yanagi, Hideki**
  Takarazuka-shi, Hyogo-ken (JP)
- **Matsumoto, Masayasu**
  Mino-shi, Osaka (JP)
- **Yura, Takashi**
  Kyoto-shi, Kyoto (JP)

(74) Representative:
VOSSIUS & PARTNER
Postfach 86 07 67
81634 München (DE)

(54) Stress proteins

(57) Described is a stress protein named ORP150, polynucleotides encoding said protein as well as antibodies against the ORP150 protein. Furthermore, pharmaceutical compositions comprising these proteins, polynucleotides or antibodies are described and their use for the treatment of ischemic diseases.

EP 0 780 472 A3

Printed by Xerox (UK) Business Services
2.16.3/3.4

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 12 0662

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	DATABASE EMROD EMBL Entry CG34206, Genbank Acc. No. U34206, 28 September 1995 CHEN, X. ET AL.: "Cricetulus griseus 170 kDa glucose regulated protein (grp170) mRNA, complete cds." XP002060254 * the whole document *	1-11	C12N15/12 C07K14/435 C12N1/21 C12N15/70 C07K16/18 A61K31/70 C12Q1/68 A61K39/00 G01N33/577
P,X	-& CHEN, X. ET AL.: "The 170 kDa glucose regulated stress protein is a large HSP70-, HSP110-like protein of the endoplasmic reticulum." FEBS LETTERS, vol. 380, no. 1-2, 12 February 1996, pages 68-72, XP002060249 ---	1-11	C12N15/79
X	NAVED, A.F. ET AL.: "CBP-140, a novel endoplasmic reticulum resident Ca(2+)-binding protein with a carboxy-terminal NDEL sequence showed partial homology with 70-kDa heat shock protein (Hsp70)." CELL STRUCTURE AND FUNCTION, vol. 20, April 1995, pages 133-141, XP002060250 * the whole document * ---	1-11	
X	EP 0 683 230 A (CALIFORNIA INST OF TECHN) * the whole document * --- -/--	15	

TECHNICAL FIELDS SEARCHED (Int.Cl.6)
C12N
C07K
A61K
G01N
C12Q

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 March 1998	Smalt, R

CATEGORY OF CITED DOCUMENTS

X : particularly relevent if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disolosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03 82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 12 0662

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	DATABASE EMROD EMBL Entry RNU41853, Genbank Acc. No. U41853, 10 August 1996 KUWABARA, K. ET AL.: "Rattus norvegicus 150 kDa oxygen regulated protein (ORP150) mRNA, complete cds." XP002060255 * the whole document *	1-11	
D	-& KUWABARA, K. ET AL.: "Purification and characterization of a novel stress protein, the 150-kDa oxygen-regulated protein (ORP150), from cultured rat astrocytes and its expression in ischemic mouse brain." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 9, 1 March 1996, pages 5025-5032, XP002060251 * the whole document *		

P,X	DATABASE EMEST10 EMBL Entry HSW18185, Genbank Acc. No. W18185, 8 May 1996 LIEUALLEN, K. ET AL.: "IMAGE:20073 Soares infant brain 1NIB Homo sapiens cDNA clone 20073." XP002060256 * the whole document *	1-11	TECHNICAL FIELDS SEARCHED (Int.Cl.6)

	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 March 1998	Smalt, R

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 12 0662

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	TSUKAMOTO, Y. ET AL.: "150-kDa oxygen-regulated protein is expressed in human artherosclerotic plaques and allows mononuclear phagocytes to withstand cellular stress on exposure to hypoxia and modified low density lipoprotein" JOURNAL OF CLINICAL INVESTIGATIONS, vol. 98, no. 8, 8 October 1996, pages 1930-1941, XP002060252 * the whole document * ---	1-4,9-14	
A	HEACOCK, C.S. ET AL.: "Enhanced synthesis of stree proteins caused by hypoxia an drelation to altered cell growth and metabolism." BRITISH JOURNAL OF CANCER, vol. 62, no. 2, August 1990, pages 217-225, XP002060253 * the whole document * -----	9	

TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 March 1998	Smalt, R

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

Application Number
EP 96 12 0662

CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claim(s):

☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

SEE SHEET B
(in case of Lack of Unity)

☒ ~~All further search fees have been paid within the fixed time limit.~~ The present European search report has been drawn up for all claims.

☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:

☐ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:

European Patent Office

LACK OF UNITY OF INVENTION
SHEET B

Application Number
EP 96 12 0662

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. Claims: 1-14 partially, and 15.

A human hypoxia-inducible protein of approx. 150 kDa, DNA encoding it, vector comprising said DNA, host cell transformed with said vector, process for preparation of the peptide by expression in said host, an antibody or fragment thereof against the peptide, a nucleic acid hybridizing to said DNA, and pharmaceutical or diagnostic preparations comprizing the DNA, peptide, antibody or hybridizing nucleic acid. Also an hypoxia-inducible promoter sequence.

2. Claims: 1-14 partially

A rat hypoxia-inducible protein of approx. 150 kDa, DNA encoding it, vector comprising said DNA, host cell transformed with said vector, process for preparation of the peptide by expression in said host, an antibody or fragment thereof against the peptide, a nucleic acid hybridizing to said DNA, and pharmaceutical or diagnostic preparations comprizing the DNA, peptide, antibody or hybridizing nucleic acid.